AD-A115 544 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 12/1
A MONTE CARLO TECHNIQUE USING LINEAR INTERPOLATION TO GENERATE --ETC(U)
DEC 81 D R ROGERS
UNCLASSIFIED AFIT/GOR/MA/81D-11 NL

1 of 2
AD A
115544

UNITED STATES AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

A MONTE CARLO TECHNIQUE USING
LINEAR INTERPOLATION TO GENERATE
MODIFIED KOLMOGOROV-SMIRNOV
CRITICAL VALUES FOR THE
EXTREME VALUE DISTRIBUTION

THESIS

AFIT/GOR/MA/81D-11 Douglas Rogers
2Lt USAF

Approved for public release; distribution unlimited.

AFIT/GOR/MA/81D-11

A MONTE CARLO TECHNIQUE USING LINEAR INTERPOLATION TO GENERATE MODIFIED KOLMOGOROV-SMIRNOV CRITICAL VALUES FOR THE EXTREME VALUE DISTRIBUTION

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

by

Douglas R. Rogers
2Lt USAF
Graduate Operations Research

December 1981

Approved for public release; distribution unlimited.

Preface

This thesis investigated the merits of an estimation technique which would supposedly reduce the number of samples necessary to obtain useful Kolmogorov-Smirnov critical values when the location and scale parameters of the hypothesized distribution were unknown. The distribution used in this investigation was the Type I, extreme value distribution (largest extreme value).

I would like to thank my advisor, Capt. Brian Woodruff, for numerous suggestions which aided in the writing of this thesis.

I would further like to thank Lt Col James Dunne who was a reader for this thesis. He gave valuable suggestions concerning the writing of this work.

In addition, I wish to thank Dr. Albert H. Moore who provided the impetus for this thesis and who offered many useful ideas.

Finally, I would very much like to thank my wife Karen whose help in preparing this document was invaluable.

Contents

List of Figures

Figure	Page

List of Tables

AFIT/GOR/MA/81D-11

Abstract

An investigation was conducted to examine the merits of an estimation technique involving linear interpolation to estimate Kolmogorov-Smirnov (K-S) critical values when the scale and location parameters of the hypothesized distribution are unknown. The purpose of the linear estimation technique is to reduce the number of Monte Carlo generated samples necessary to produce useful critical values for the K-S goodness-of-fit test.

Also, different plotting positions were studied to ascertain which plotting positions used in calculating and plotting the K-S test statistic values provided the best critical value estimation.

The distribution used as the hypothesized distribution was the Type I, extreme value distribution (largest extreme value).

In addition, a power study was performed which compared the power of the true critical values against the power of the estimated critical values.

The following are the major results. Useful critical values were found with the linear estimation technique using relatively few Monte Carlo generated samples.

Further, the plotting positions found to be the best in calculating the K-S test statistic values and plotting these values were respectively i/n and $(i-.5)/n$ where i is the ith ranked point in a sample of size n.

A MONTE CARLO TECHNIQUE USING LINEAR INTERPOLATION TO GENERATE MODIFIED KOLMOGOROV-SMIRNOV CRITICAL VALUES FOR THE EXTREME VALUE DISTRIBUTION

I. Introduction

The extreme value distribution has applications in many fields; however, its application to reliability (Ref 12), particularly failure due to corrosion, and meteorology (Ref 11), make it important to the Air Force. Because aircraft reliability and meteorological phenomena is of vital interest to the Air Force, data pertaining to both areas is collected by the Air Force. It is usually useful to determine the underlying distribution of the reliability or meteorological data. A "goodness-of-fit" test can be performed to determine the underlying distribution for the data.

A goodness-of-fit test is essentially a comparison between a hypothesized distribution and the frequency distribution of the sample data. If the sample data "fits" the hypothesized distribution adequately, then the hypothesized distribution is used as the representative for the underlying distribution.

The Kolmogorov Goodness-of-fit Test

The Kolmogorov goodness-of-fit test, which will be described, is often referred to as the Kolmogorov-Smirnov

one-sample test. In this work, the test shall be referred to merely as the Kolmogorov-Smirnov (K-S) test.

The purpose of the K-S test is to determine if a hypothesized, completely specified, continuous distribution $F_h(x)$, is the underlying distribution of a set of sample data, $(x_1, x_2, \ldots, x_n)$. This is carried out by comparing the cumulative distribution function $F_h(x)$ of the hypothesized distribution to the cumulative step function $S_n(x) = i/n$; i equals the number of data points less than or equal to x, and n is the number of data points in the sample. If $F_h(x)$, the proportion of the population defined by the hypothesized distribution having values less than x, is adequately close to the step function $S_n(x)$, then the hypothesized distribution is accepted as the underlying distribution. A graphical view of this comparison is presented in Figure 1.

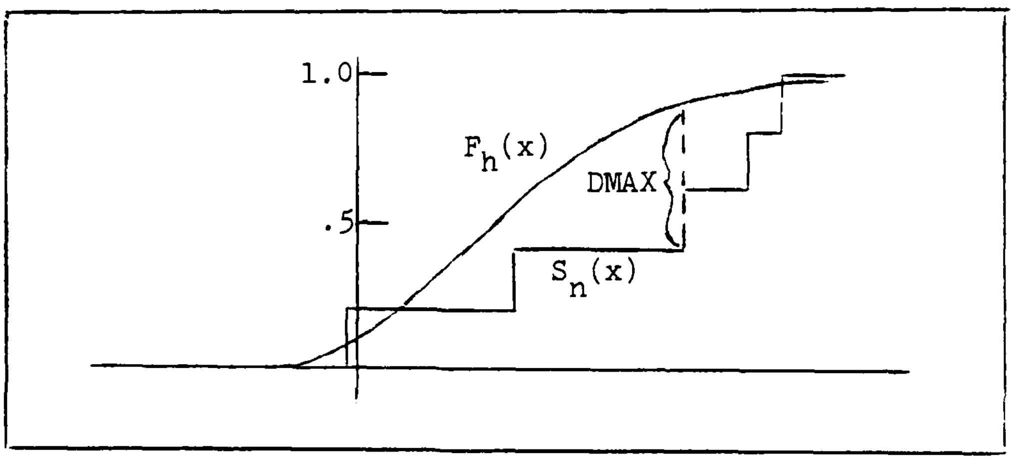

Fig 1. Graphical View of the K-S Test (Ref 4:346)

One determines if $S_n(x)$ and $F_h(x)$ are adequately close by using the maximum vertical distance between $F_h(x)$ and $S_n(x)$ as a test statistic. DMAX represents this maximum vertical distance in Figure 1. In mathematical terms, the maximum vertical distance is

$$\text{DMAX} = \text{MAX}\left|F_h(x) - S_n(x)\right| \tag{1}$$

If the data used to calculate DMAX is from the hypothesized distribution, the distribution of DMAX is known and is independent of $F_h(x)$; however, the distribution is dependent on n (Ref 2). For various values of n, certain DMAX values are found in standard tables of critical values (Ref 20:70). If the DMAX statistic for a sample of size n is larger than the appropriate critical value, the hypothesized distribution is rejected as being the underlying distribution.

Problem

The standard tables of critical values for the K-S test can only be used when $F_h(x)$ is completely specified. In other words, all parameters associated with $F_h(x)$ have known values. However, the probability integral transformation used by David and Johnson (Ref 8) allows critical value tables for the K-S test to be made for any particular distribution when the scale and location

parameters of $F_h(x)$ are unknown and estimated from the sample data. The distribution of DMAX, or any test statistic based on the cumulative distribution function, is independent of the scale and location parameters, but it is dependent on the functional form of the hypothesized distribution, $F_h(x)$ (Ref 8:190). Thus, when either the scale or location parameters, or both, must be estimated, the standard critical value tables cannot be used, and new critical values must be found for each hypothesized distribution. Lilliefors has tabulated critical values for the normal distribution with unknown means and variance (Ref 17). Lilliefors has also generated tables for the exponential distribution with the mean unknown (Ref 18). Tables of critical values have also been produced for the three-parameter Weibull distribution and the gamma distribution where the scale and location parameters in each were estimated (Ref 7).

The cumulative distribution function of the extreme value distribution, which will be defined fully later, has associated with it only two parameters. These parameters are scale and location parameters. Hence, critical value tables can be generated for the extreme value distribution with unknown parameter values. However, this table can only be created through Monte

Carlo simulation (computer generation of an empirical distribution of DMAX values) which requires a large amount of computer time.

The large amount of computer time is necessary to generate a sample size of DMAX values large enough to create an empirical distribution which will yield valid critical values. One of the research questions of this study is: Is there a technique whereby the necessary sample size of DMAX values can be lessened?

Further, the cumulative step-function $S_n(x)$ defined earlier is also referred to as a plotting position or a plotting convention. It is the plotting position normally used in Eq (1) to calculate DMAX. It is also the plotting position usually utilized in plotting the DMAX values in order to obtain an empirical distribution for the DMAX statistic. A second research question is: Is $S_n(x)$ the best possible plotting position to use in a technique which would lessen the necessary number of DMAX values created through Monte Carlo simulation?

Background

Some preliminary work has been done that may aid in answering the questions put forth in the previous two paragraphs.

Johnston (Ref 14), working with component reliabilities, as opposed to defining critical values of a statistic, found good results from empirical distributions formed from small, Monte Carlo-generated samples, linear interpolation, and linear extrapolation. In other words, linear interpolation between the data points vaguely outlining the empirical distribution and extrapolation to estimate endpoints allowed a small number of data points to be used to create a usable representation of the true distribution of the component reliabilities. A small number of data points means a small amount of computer time.

Also, other plotting positions have been proposed (Ref 1:145; 15:300-301; and 16:548-551) as better estimates of cumulative distribution functions than $S_n(x)$. One of the other plotting positions may yield a K-S table of critical values, created by the estimating technique previously discussed, that is more powerful than a table derived using $S_n(x)$.

Objectives

This work has three objectives. The first objective is to discover if linear interpolation between the Monte Carlo-generated data points (DMAX values) outlining the DMAX empirical distribution and extrapolation

used to obtain endpoints will reduce the DMAX sample size required to find suitable critical values.

The second objective is to investigate the possibility that plotting positions other than $S_n(x)$ can produce better critical value estimates.

The third and final objective, which is a product of the first two objectives, is to fashion a table of critical values for the estimating technique discussed.

Overview

This study provides a more detailed discussion on several of the points brought out in this chapter. Particular attention is given to the description and application of the extreme value distribution in Chapter II.

Chapter III gives a presentation of the standard technique for finding critical values along with a description of an estimation technique. Also, in Chapter III, plotting positions are discussed.

Chapter IV essentially describes those activities undertaken to collect and analyze the data used in this study.

Chapter V tells by way of an example how to use critical value tables for the K-S test.

Chapter VI presents the research results.

Chapter VII provides concluding remarks about this study and the topic area it covered. Also, comments concerning how this study could have been improved and how this study could be expanded are offered in Chapter VII.

Finally, the appendices present the reader with tables and figures which are useful in understanding the development and results of this thesis.

II. The Extreme Value Distribution

Many times when doing a statistical analysis, the major interest concerns the largest or smallest extreme values. Thus, a need exists to understand the limiting form of the distribution of the smallest or largest extreme value. It was from this need that the extreme value distribution arose.

Evolution

Consider a sample consisting of n data points where the values are arranged in ascending or descending order. Let each data point x be given a subscript i so that i indicates order. If arranged in ascending order, x_1 is the smallest value and x_n is the largest. If several ordered samples are under study and interest is directed to one of the extremes in each sample, then it may be practical to have some knowledge of the probability distribution which applies to such a set of extremes.

Consider the set of data $(x_1, x_2, \ldots, x_n)$ in which the elements are in no particular order, and each element is defined by the probability density function g(x) or the cumulative distribution function G(x). Also, let W be defined as

$$W = \max(x_1, x_2, \ldots, x_n) \tag{2}$$

Then, the cumulative distribution function of W, in terms of x, can be defined as H(x) which leads to

$$\begin{aligned} H(x) &= P(W \leq x) \\ &= (x_1 \leq x, x_2 \leq x, \ldots, x_n \leq x) \end{aligned} \tag{3}$$

Assuming all x_i's are independent, Eq (3) becomes

$$\begin{aligned} H(x) &= \prod_{i=1}^{n} P(x_i \leq x) \\ &= [G(x)]^n \end{aligned} \tag{4}$$

The probability density function h(x) is then the derivative of Eq (4) or

$$h(x) = n[G(x)]^{n-1}\ g(x) \tag{5}$$

Now, consider

$$\begin{aligned} \text{Ln } H(x) &= \text{Ln } [G(x)]^n \\ &= n\text{Ln } G(x) \end{aligned} \tag{6}$$

which leads to

$$H(x) = \exp[n\text{Ln } G(x)]$$

For large values of n, the exponent of e in Eq (7) approaches the indefinite form infinity times zero.

Thus, the limiting (or asymptotic) distribution of the largest extreme value is dependent on the manner in which G(x) approaches unity (Ref 23:66).

Johnson and Kotz (Ref 13:275-276), explaining Gnedenko's original results, show the correspondence between the behavior of G(x) as x approaches infinity and the Type extreme value distribution to which the limiting distribution belongs. Three Types of extreme value distributions exist.

The condition which leads to the Type I extreme value distribution for the largest extreme value is, as x approaches infinity, the limit of (1 - G(x)) converges to zero as quickly as the limit of exp(-x) (Ref 23:66). Several parent distributions, the distribution defined by G(x), meet this condition; some of them are the normal, log-normal, exponential, logistic, and gamma distributions (Ref 9:5). It is the Type I extreme value distribution for the largest extreme value with which this work is concerned.

The condition leading to the Type I extreme value distribution for the smallest extreme value is, as x approaches infinity, the limit of G(x) converges to zero as quickly as the limit of exp(-x) (Ref 15:43).

To give a broader coverage of the extreme value distribution it will be noted that the Type II limiting distribution is a limiting form of the parent distribu-

tion of the Pareto or Cauchy type. The Type II extreme value distribution has not been applied as extensively as the Type I or Type III forms and is considered to be of lesser importance than the other forms. Further, the Type III form arises when the range of the density function defining the parent distribution is bounded from below ($x \geq u$), and the cumulative distribution function of the parent distribution behaves similarly to $(b-u)^a$ as x approaches u for some $b>0$ and $a>0$ (Ref 15:153). The Type III extreme value distribution for the smallest extreme value is the often used Weibull distribution, which has been applied to metal fatigue and fracture problems.

Description

For brevity and since Type I is the one most commonly referred to as the "extreme value distribution," the term "extreme value distribution" will be, from this point forward, in reference to the Type I extreme value distribution.

The cumulative distribution function of the extreme value distribution (largest extreme value) (Ref 13:272) is

$$F(x) = \exp[-\exp(-(x-u)/b)] \tag{8}$$

and in the same manner the progression was made from Eq (4) to Eq (5), the probability density function is (Ref 22:9)

$$f(x) = b^{-1}\exp[-(x-u)/b]\exp[\exp(-(x-u)/b)] \qquad (9)$$

where for Eqs (8) and (9) $-\infty < X < \infty$.

The cumulative distribution function and the probability density function for the extreme value distribution (smallest extreme value) are respectively (Refs 15:43; 22:9)

$$F(x) = 1 - \exp[-\exp(x-u)/b] \qquad (10)$$

and

$$f(x) = b^{-1}\exp[(x-u)/b]\exp[-\exp(x-u)/b)] \qquad (11)$$

where for Eqs (10) and (11) $-\infty < X < \infty$.

The distributions defined by Eqs (8), (9), (10), and (11) have a scale parameter b and a location parameter u; no shape parameter is involved. The graphic representation of Eqs (9) and (11) can be seen in Figure 2. Each is unimodal with its mode at u and inflection points at (Ref 13:278)

$$u \pm b\mathrm{Ln}\,[\tfrac{1}{2}(3+(5)^{\frac{1}{2}})] \qquad (12)$$

Actually, there are several symmetry relationships between the extreme value distribution (largest extreme value) and the extreme value distribution (smallest extreme value). These are seen in Table I and Figure 2. As can be seen in Figure 2, both distributions are skewed yet mutually symmetrical about u.

Applications

The theoretical basis of the extreme value distribution involves consideration of the parent distribution; however, application of the extreme value distribution does not always demand that the parent distribution be identified. In some cases, the extreme value distribution has been applied effectively when no theoretical support existed for its use. This is not to say that it should be applied indiscriminately, only that it is not always necessary to analyze the nature of the parent distribution. This has allowed the extreme value distribution (largest extreme value) to be utilized in many diverse areas.

Gumbel (Ref 11) has applied it to the analysis of structural fatigue. Gumbel has also shown applications of the extreme value distribution (largest extreme value) to flood flows, droughts, and other extreme meteorological phenomena (Ref 12) as well as communication systems

Table I

Characteristics of Extreme Value Distribution (Ref 22:11)

Characteristic	Smallest Value	Largest Value
Mean	u - Sb	u + Sb
Standard Dev.	$b\,\pi/\sqrt{6}$	$b\,\pi/\sqrt{6}$
Median	u - .36651 b	u + .36651 b
Mode	u	u

S = .57721...

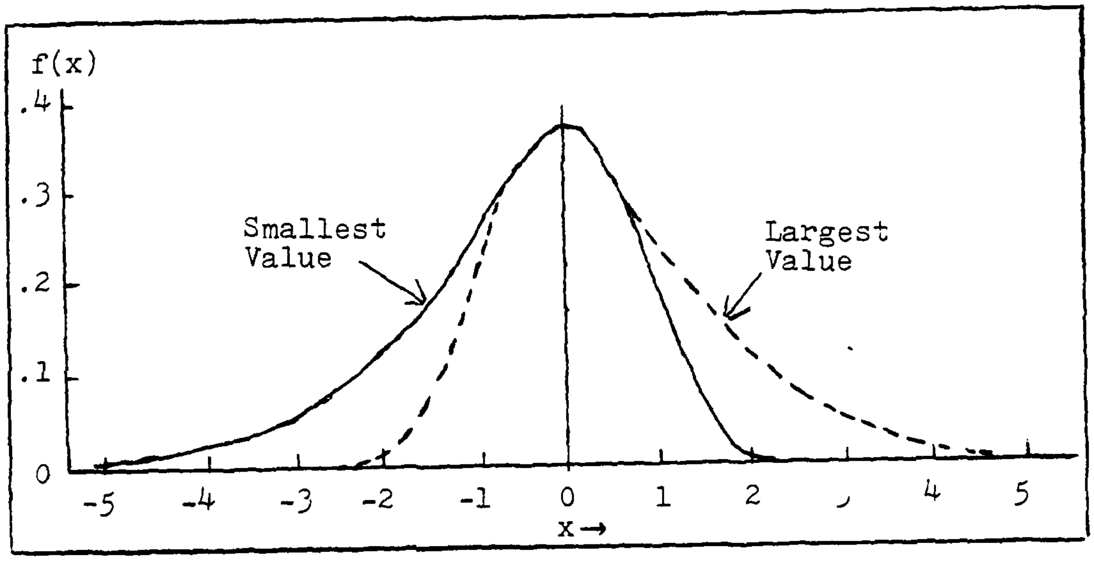

Fig 2.

Extreme Value Distributions of Smallest and Largest Values with b = 1, u = 0 (Ref 22:11)

(Ref 10). Bogdanoff and Schiff (Ref 3:147-178) have shown the applicability of the distribution to the prediction of the force of earthquakes. Johnson and Kotz (Ref 13:274) have referred to numerous applications; microorganism survival times, corrosion, and dielectric strengths of paper capacitors. Kapur and Lamberson cite its use in analyzing gust velocities, gust loads, and landing loads for aircraft (Ref 15:154). It has been used by Branger (Ref 4:213-237) to estimate the life of fighter aircraft.

Chapter II Summary

A need to analyze the extreme values of a sample was the impetus in the evolution of the extreme value distribution. In the early theoretical work on the extreme value distribution, it was discovered that the extreme value distribution has three forms. These forms are called Type I, Type II, and Type III. The Type of the extreme value distribution is dependent on the limiting properties of the parent distribution.

The extreme value distribution is applied widely. It is even being applied successfully in circumstances where the theoretical foundation for its use is weak. Several of the areas in which the extreme value distribution is applied are structural fatigue, flood flows, and gust velocities.

III. The Techniques and Plotting Positions

The standard method of finding critical values for the K-S test when the hypothesized distribution is fully specified usually involves a Monte Carlo technique which generates a large number of samples. Normally, 5,000 to 10,000 samples are generated. From each of these samples a DMAX value is found, thus giving a large sample of DMAX values. With such a large number of DMAX values, an excellent approximation to the cumulative distribution function of the DMAX values can be made. Actually, with so many values, the critical values found using this standard technique are taken as truth.

On the other hand, the linear estimation technique uses a drastically reduced number of samples as compared to the standard technique.

Additionally, it should be pointed out that plotting positions other than $S_n(x)$ have been advanced. The use of an alternative plotting position may offer a better linear estimation technique.

Standard Technique

The method which shall be discussed is the method normally used to obtain critical values at their corresponding significance levels for the K-S test when the hypothesized distribution is not fully specified.

For a fixed sample size n, a sample of extreme value (largest extreme value) random deviates is generated. The scale and location parameters are then estimated since the hypothesized distribution is not fully specified. DMAX, which was introduced in Eq (1), is calculated for the sample. This procedure is repeated using a different set of sample elements each time. Assume the procedure is repeated 5,000 times thus giving a sample consisting of 5,000 independent DMAX values. The 5,000 DMAX values are ranked and the 80th, 85th, 90th, 95th, and 99th percentiles are found. The percentiles are interpreted then as the critical values.

Ranking the DMAX values and finding the percentile values is analogous, in a graphical sense, to plotting the DMAX values with respect to the vertical axis using $S_n(x)$. Thus, the vast number of plotted points is an approximation to the plot of the cumulative distribution function of the DMAX values. This interpretation can be seen in Figure 3. With 5,000 points, the curve is, for all practical purposes, completely defined.

It should be stated that associated with each critical value is a significance level. The significance level, often denoted merely by α, is the risk of rejecting the hypothesized distribution as the true underlying distribution when in fact, it is the underlying distribu-

tion. In other words, the DMAX value for a sample of size n which has the extreme value distribution (largest extreme value) as its underlying distribution will be greater than the critical value (denoted $DMAX_{n}, \alpha$) associated with sample size n and significance level α with probability α. Thus, the critical values corresponding to the 80th, 85th, 90th, and 99th percentiles have significance levels which are respectively .20, .15, .10, .05, and .01.

The Estimation Technique

The technique for estimating critical values is described below. It is detailed in terms of a graphical procedure for reasons of clarity.

Initially, for a fixed sample of size n, a sample of extreme value (largest extreme value) random deviates is generated, and from these n random deviates, the scale and location parameters are estimated. Using these estimated parameter values, DMAX is calculated for the sample. This procedure is repeated N times with a different sample of random deviates for each repetition. The N DMAX values are plotted with respect to the vertical axis using a plotting position, and lines are then drawn between the plotted points. This concept can be seen in Figure 4.

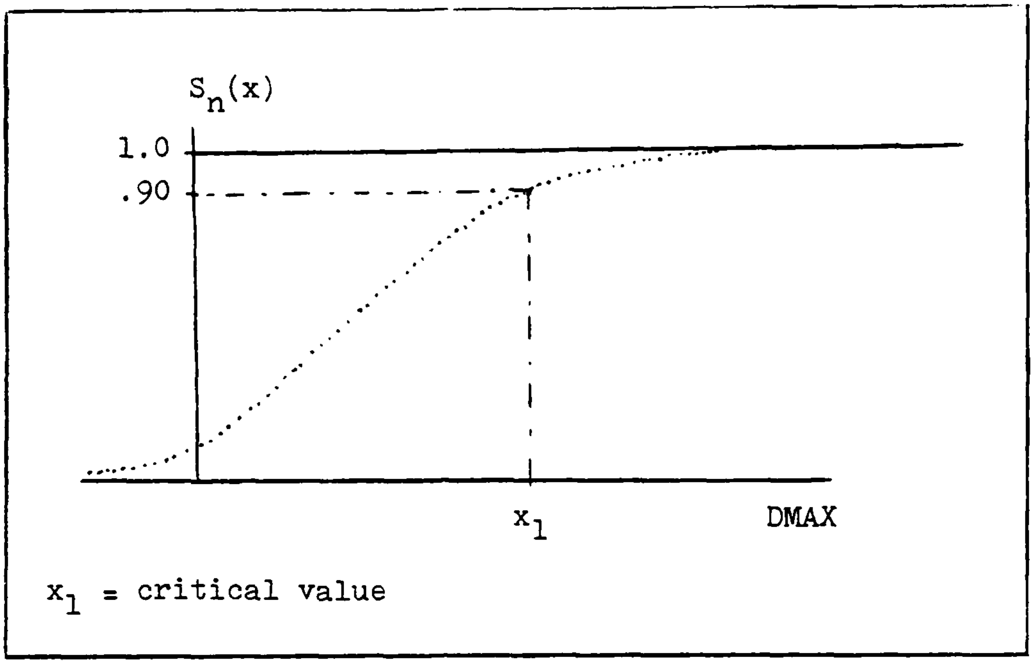

Fig 3. Graphical Representation of Standard Technique

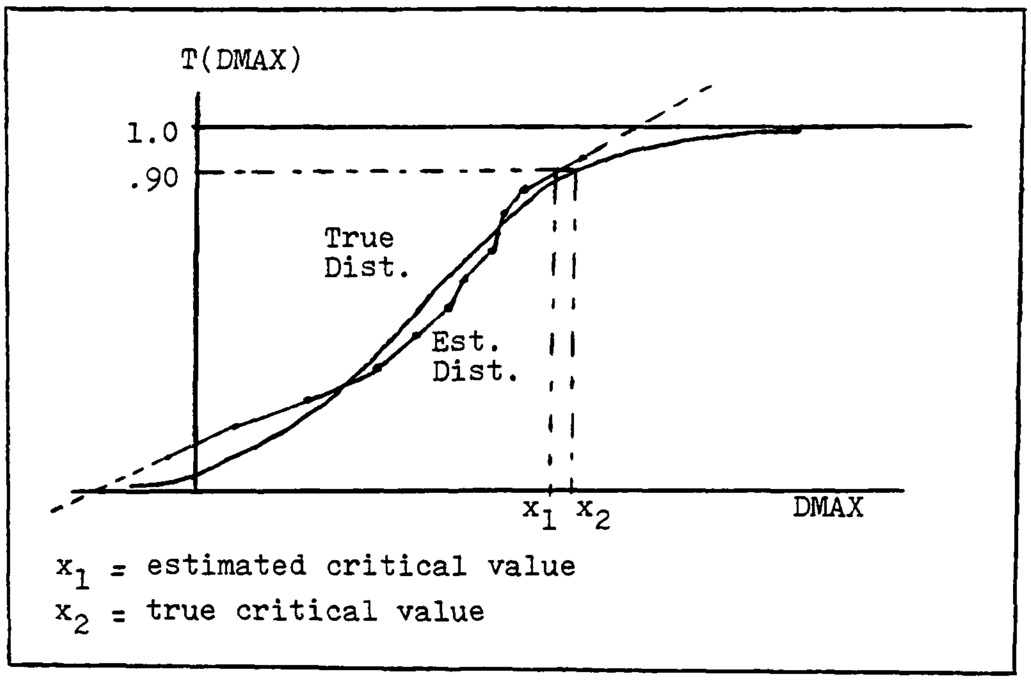

Fig 4. Graphical Representation of Linear Estimation Technique

Further, lines are extrapolated outward to create endpoints; this can be seen in Figure 4 where the segmented lines represent the extrapolated lines. The endpoints are formed at T(DMAX) = 1 and T(DMAX) = 0 where T(DMAX) is the cumulative distribution function of DMAX.

The critical value $DMAX_{n,\alpha}$ is found by discerning what DMAX value corresponds with the vertical axis value $1 - \alpha$. For example, the critical value $DMAX_{n,.10}$ is found in Figure 4 by deducing the DMAX value which pairs with F(DMAX) = .90. All critical values are found in this manner. As can be seen in Figure 4, the bar-dot lines which "pick-off" the critical value intersect not at one of many points, as in Figure 3, but at the line drawn between two of merely a few points; thus, the critical value is estimated.

Plotting Positions

In the description of the standard technique for finding critical values, it was alluded that the plotting position $S_n(x)$ is used as an approximation to $F_h(x)$ and thus is used in the calculation of DMAX. And, as already mentioned $S_n(x)$ is used to "plot" the DMAX values. However $S_n(x)$ is not the only plausible plotting position. An alternative may prove to be more useful in the linear

estimation technique. Blom, Kimball, and others have suggested alternate plotting positions with which to use to approximate cumulative distribution functions (Refs 1:145; 15:301, and 16:548-549).

Five alternative plotting positions are investigated in this work. One of the alternatives investigated is

$$F(x_i) \approx \frac{i}{n+1} \tag{13}$$

It can be shown that Eq (13) is the mean value of the fraction of the population, defined by the cumulative distribution function F(x), falling prior to the ith ranked data point in a sample of size n (Ref:279-299). Eq (13) is often called the mean rank.

Another alternate plotting position is

$$F(x_i) \approx \frac{i-.3}{n+.4} \tag{14}$$

Eq (14), called the median rank, is the median value of the fraction of the population, defined by F(x), coming before the ith ranked data point.

The other plotting positions examined are

$$F(x_i) \approx \frac{i-.5}{n} \tag{15}$$

$$F(x_i) \approx \frac{i-.375}{n+.25} \tag{16}$$

$$F(x_i) \approx \frac{\frac{i}{n+1} + \frac{i-1}{n-1}}{2} \tag{17}$$

Chapter III Summary

The standard technique for finding critical values requires that a large number of DMAX values be generated. Normally, 5,000 to 10,000 DMAX values are used. This large number of DMAX values "plotted" using $S_n(x)$ defines the DMAX cumulative distribution so closely that the critical values found using this technique are considered true.

The linear estimator technique uses drastically fewer DMAX values. These DMAX values are plotted using a plotting position, not necessarily $S_n(x)$. Lines are drawn between the plotted DMAX values, thus giving a rough outline of the graph of the cumulative distribution of the DMAX values.

Several different plotting positions have been suggested as alternatives to $S_n(x)$. At least one of these, as opposed to $S_n(x)$, may supply a better linear estimating technique.

IV. Methodology

The following sections describe the overall process that was undertaken to acquire and analyze the data needed for this study. A detailed description is given, and then, in the summary section a step-by-step overview of the methodology process used in this work is presented.

Initial Process

Initially, the standard technique, using 5,000 DMAX values, was employed to generate critical values for $n = 10$, 25, and 40 at the significance levels .20, .15, .10, .05, and .01. These critical values, which shall be referred to as the standard set of critical values, were taken as the true critical values and used for comparison with the estimated critical values.

Next N DMAX values were generated at $n = 10$, 25, and 40. These were utilized in the linear estimation technique to estimate the true critical values at the significance levels .20, .15, .10, .05, and .01. The N DMAX values were calculated using $S_n(x)$ in Eq (1) and plotted using $S_n(x)$ and each of the five alternative plotting positions thus giving six sets of estimated critical values. A set of estimated critical values was compared to the standard set of critical values by calculating the mean absolute difference between the

elements of the estimated set and the corresponding elements of the standard set. An additional comparison was made for each estimated set; note was taken of the largest absolute difference between the elements of the estimated set and the corresponding elements in the standard set. This comparison procedure involving the mean absolute difference and the largest absolute difference occured for N = 100, 250, 500, and 750.

It was viewed that a "good" estimation would be approximately a mean absolute difference of .002 and a largest absolute difference of .004. It should be stated that the two values given as measures of a "good" estimation were arbitrarily selected. It was also felt that going to N larger than 750 to reach the measures of a "good" estimation would signify that the linear estimation technique is a poor estimating technique. This view was taken since the primary objective of the technique is to markedly reduce N, the number of DMAX values.

From comparing each of the six sets of estimated critical values to the standard set, the best plotting position with which to plot the DMAX values was chosen. This plotting position was then used to plot the DMAX values in all additional implementations of the linear estimation technique.

Of the five alternative plotting positions to $S_n(x)$, the three best in terms of plotting the DMAX values were chosen to be analyzed as alternatives to $S_n(x)$ in the calculation of DMAX. It was felt that the plotting positions which best estimated the cumulative distribution function of DMAX might provide the best estimation of $F_h(x)$, in this case, the cumulative distribution function of the extreme value distribution (largest estreme value). Thus, three new sets of estimated critical values were compared to the standard set giving a mean absolute difference value and a largest absolute difference value for each estimated set. From these values it was determined if substituting $S_n(x)$ with some other plotting position from the five alternatives set forth would provide a better linear estimation technique.

Essentially, the proceeding process gave analysis in three important areas of this study. First, it analyzed the matter of whether or not the linear estimation technique is a "good" technique in terms of providing critical values "close" to the true critical values. Secondly, it investigated other plotting positions as substitutes for $S_n(x)$ in plotting the DMAX values. Finally, it studied the possibility that at least one of the alternative plotting positions should replace $S_n(x)$ in the calculation of DMAX.

One further topic was analyzed; that was the power of the linear estimation technique. How this analysis was done is discussed in the following section.

Power Study

The term power is defined as the probability of rejecting the hypothesized distribution as the underlying distribution of a sample when the hypothesized distribution is not the underlying distribution. From this definition, it can be seen that a goodness-of-fit test with high power values is desirable.

Initially, a power study was conducted on the linear estimation technique (with N = 100, 250, 500, and 750) formed from the plotting positions that were determined to give the best technique.

The power study on the linear estimation technique was executed by generating 2,500 samples of some distribution other than the extreme value distribution (largest extreme value). For each sample, a DMAX value was calculated using the plotting position which was chosen earlier to be the best in calculating DMAX. These DMAX values were compared to the appropriate critical values, and the number of rejections of the extreme value distribution (largest extreme value) as the underlying distribution were counted. The proportion of rejections out of the 2,500 trials was taken as the power.

This process was conducted for the following distributions: the two-parameter Weibull (shape = 1, scale = 1), the standard normal, the two-parameter log-normal (mean = 0, standard deviation = 1), the chi-square (1 degree of freedom), and the chi-square (4 degrees of freedom). Power values were also found using the same distributions, sample sizes, and significance levels for the standard critical values. Using $S_n(x)$ in the calculation of DMAX, 2,500 DMAX values were found and compared with the appropriate standard critical values. The power values from the estimated technique were compared to the power values from the standard technique.

Generation of Random Deviates

In this study it was necessary to generate random deviates from six different distributions. These distributions were the extreme value (largest extreme value), the two-parameter Weibull, the standard normal, the two-parameter log-normal, the chi-square (1 degree of freedom), and the chi-square (4 degrees of freedom).

The extreme value (largest extreme value) random deviates were generated with the use of two-parameter Weibull random deviates. The two-parameter Weibull distribution is defined by the cumulative distribution function

$$F(t) = 1 - \exp[-(t/z)^m], \quad t \geq 0 \tag{18}$$

where the shape parameter m and the scale parameter z are both assumed to be positive.

If the transformation $x = -Ln(t)$ is made when t has the distribution defined by Eq (18), then x is distributed according to the extreme value distribution (largest extreme value) (Ref 5:2). The scale and location parameters of the extreme value distribution (largest extreme value) are then $b = m^{-1}$ and $u = -Ln(z)$ respectively.

Thus, in order to generate extreme value random deviates, two-parameter Weibull random deviates were generated. The parameter values used were $z = 1$ and $m = 1$. The Weibull random deviates were generated on the CDC 6600 computer by the International Mathematical and Statistics Library (IMSL) subrouting GGWIB. The transformation $x = -Ln(t)$ was performed on the two-parameter Weibull random deviates, and extreme value random deviates with $u = 0$ and $b = 1$ were then produced.

The standard normal random deviates were generated by using the IMSL subroutine GGNML.

The two-parameter log-normal random deviates were generated with the use of the subroutine GGNLG which is also from IMSL.

The chi-square (1 degree of freedom) random deviates were generated through the use of standard normal random deviates. It can be shown that the square of a standard

normal random variable has a chi-square (1 degree of freedom) distribution (Ref 21:217-218). Using that fact, standard normal random deviates were generated, and these random deviates were squared. Thus, chi-squared (1 degree of freedom) random deviates were formed.

The chi-square (4 degrees of freedom) random deviates were also generated using standard normal random deviates. The method that was employed is the same as that used by Littell, McClave, and Offen (Ref 19:265). Each chi-square (4 degrees of freedom) random deviate was constructed by summing the squares of four independent standard normal random deviates.

Estimation of Parameters

The scale and location parameters were estimated by using recursive equations presented by Stephens (Ref 24). These equations provide maximum likelihood of the parameters. The equation for the location parameter is

$$u = -b \operatorname{Ln} \left[n^{-1} \sum_j \exp(-x_j/b) \right] \tag{19}$$

and the equation for the scale parameter is

$$b = n^{-1} \sum_j x_j - \left[\sum_j x_j \exp(-x/b) \right] \left[\sum_j \exp(-x/b) \right]^{-1} \tag{20}$$

In both Eqs (18) and (19), n is the number of elements in the sample, and x_j is the jth element in the sample. Further, all summations represent sums from j = 1 to n.

As can be seen in the equations, b must be solved for recursively to some arbitrary tolerance level and then substituted into Eq (18) to find u. The tolerance level used in finding b was .00001. That is, if the change in the previous estimate of b is less than or equal to .00001 when compared to the current estimate, the estimation process is discontinued, and the current estimate is accepted. The maximum number of iterations allowed was 500; however, this was never reached.

It should be noted that in order to initiate the recursive process to find b, an initial guess of b must be made. This guess was b = 1 in all circumstances.

Computer Programs

The computer programs utilized in this work can be found in Appendix G. These computer programs were made to be adaptable by allowing the user to easily change program parameters. In addition, the programs are segmented by comment-statement headings which state the function of the segment it designates.

Chapter IV Summary

The computer programs used in this study were made

flexible by permitting the user to quickly change program parameters.

Random deviates were needed from several different distributions. The distributions were the extreme value (largest extreme value), the standard normal, the two-parameter log-normal, the two-parameter Weibull, the chi-square (1 degree of freedom), and the chi-square (4 degrees of freedom).

The scale and location parameters were estimated by recursive equations. The tolerance level used was .00001, and the maximum number of iterations allowed was 500.

The other actions taken in the process of acquiring and analyzing data can be roughly outlined in the steps which follow:

1. A standard set of critical values was formed.

2. A determination was made as to whether or not the linear estimation technique was a "good" estimation technique.

3. A determination was made as to which plotting position best plotted the DMAX values.

4. Of the five alternative plotting positions, the three best in terms of plotting DMAX values were analyzed as substitutes for $S_n(x)$ in calculating DMAX.

5. It was determined which, if any, of the three

plotting positions chosen in step 3 should be substituted for $S_n(x)$ in DMAX calculations.

6. The plotting positions selected in step 2 and step 4 were designated as forming from the plotting positions studied the best linear estimation technique.

7. A power study was performed on the linear estimation technique found in step 5.

V. Use of Critical Value Table

The process required to use a K-S critical value table, when the location and scale parameters are estimated, is quite simple. The DMAX value is calculated for the n data points which make up the sample. This DMAX value is compared to $DMAX_{n,\alpha}$. If DMAX for the sample is greater than $DMAX_{n,\alpha}$, the hypothesized distribution is rejected as the underlying distribution of the sample. A better explanation can be provided through a demonstration.

An Example

The example which follows involves the use of the critical value table generated using the standard technique; however, the critical value table generated with the linear estimation technique is used in the same manner.

Assume the sample data -.56, 1.54, 0.17, 3.42, 2.11, 1.16, 4.92, -1.23, -1.71, 3.72, 4.23, 1.91, 3.37, -2.36, and 2.13 has been collected. Also assume that the extreme value distribution has been hypothesized as the underlying distribution of the sample and that this hypothesis is to be tested.

Before the DMAX value for the sample can be found, the scale and location parameters must be estimated from the

fifteen sample elements. These parameters could possibly be estimated by hand from Eqs (19) and (20) but only with some difficulty. Fortunately, Eikman has tabled two sets of coefficients which can be used to calculate unbiased nearly best linear estimates of the scale and location parameters of the extreme value distribution (largest extreme value) (Ref 9:32-423). By multiplying each sample element by the appropriate coefficient and summing the products, the scale or location parameter can be found depending on which set of coefficients is used. The ease of computation using the coefficients more than compensates for the slight sacrifice in accuracy of the estimate. Now, using the two sets of coefficients on the sample elements reveals estimates of 2.14 for the scale parameter and 0.34 for the location parameter.

Substituting the estimated parameter values into the cumulative distribution function of the extreme value distribution (largest extreme value) allows Table II to be constructed. In Table II, the DMAX value is prominently indicated by asterisks. From Table III, the standard table of critical values, it can be seen that $DMAX_{15,.05}$ = .2194. Since the DMAX value for the sample is less than the critical value $DMAX_{15,.05}$, the extreme value distribution (largest extreme value) cannot be rejected as the underlying distribution.

Chapter V Summary

The steps necessary to use a K-S critical value table for the extreme value distribution (largest extreme value) when the location and scale parameters are unknown are these:

1. Estimate the location and scale parameters from the sample data.

2. Find the DMAX value for the sample.

3. Compare the DMAX value for the sample to the critical value $DMAX_{n,\alpha}$.

4. If the sample DMAX value is greater than $DMAX_{n\alpha}$, reject the extreme value distribution (largest extreme value) as the underlying distribution. If the sample DMAX value is less than or equal to $DMAX_{n,\alpha}$, do not reject.

Table II

Use of Critical Value Table

i	x_i	$F(x_i)$	$S_n(x_i)$	$\lvert F(x_i)-S_n(x_i)\rvert$	$\lvert F(x_i)-S_n(x_{i-1})\rvert$
1	-2.36	.029	.067	.038	.029
2	-1.71	.074	.133	.059	.007
3	-1.23	.125	.200	.075	.008
4	-0.56	.218	.267	.049	.018
5	0.17	.338	.333	.005	.071
6	1.16	.506	.400	.106	* .173 *
7	1.54	.565	.467	.098	.165
8	1.91	.619	.533	.086	.152
9	2.11	.646	.600	.046	.113
10	2.63	.710	.667	.043	.110
11	3.37	.785	.733	.052	.118
12	3.42	.789	.800	.011	.056
13	3.72	.814	.867	.053	.014
14	4.23	.850	.933	.083	.017
15	4.92	.889	1.00	.111	.044

VI. Discussion of Results

The discussion of results which follows is begun by covering the topic of computer program validity. Once computer program validity is established, the results are then presented. Many of the results can be found in tables and figures in the appendices.

Computer Program Validity

Before a major discussion of the results begins, the question of program validity must be answered.

In order to measure the quality of the linear estimation technique as an estimator of the true critical values, it is certainly necessary that the set of critical values used as the standard set be correct. In other words, the validity of the computer program generating the standard technique critical values must be verified.

This verification was initiated by generating in the program extreme value distribution (largest extreme value) random deviates. The random deviates and the location and scale parameters were transformed to random deviates and parameters appropriate to the two-parameter Weibull distribution (see page 25). Within the program, the extreme value distribution (largest extreme value) cumulative distribution function was replaced by the cumulative distribution function for the two-parameter Weibull, and critical values were found at $n = 30$ for

significance levels .20, .15, .10, .05, and .01.

These critical values should have been two-parameter Weibull critical values. In order to verify this and thus the program validity, these critical values were compared to two-parameter critical values obtained at the same sample size and significance levels by Littell, McClave, and Offen (Ref 19:262). Table III presents the two sets of critical values. As Table III indicates, the two sets of critical values are almost identical, thus establishing the validity of the computer program.

Selection of Plotting Positions

In Appendix A is the table of critical values generated by the standard technique. It was the critical values at n = 10, 25, and 40 which formed the standard set of critical values, the set of critical values taken as truth.

Table III

Program Validity

	.20	.15	.10	.05	.01
Littell's Results	.130	.136	.144	.156	.179
Thesis Results	.130	.137	.145	.157	.181

In Appendix B are tabled the critical values generated by the linear estimation technique. The critical values in the tables in Appendix B were formed by calculating the DMAX values with the plotting position $S_n(x)$ but plotting the DMAX values with each of the six different plotting positions discussed in this study. Each different plotting position used to plot the DMAX values forms a different segment in the tables. Also, under each segment is given the mean absolute difference and the largest absolute difference related to that set of critical values. By examining the various mean and largest absolute value differences, it is seen that the plotting positions defined by Eqs (15) and (17) at N = 750 do the best job of reaching the criterion values deemed acceptable.

It should be noted that at N = 750 all the plotting positions used to plot the DMAX values offer acceptable estimates. However, Eqs (15) and (17) are equally better than the others by providing a mean absolute difference of .0012 and a largest absolute difference of .0025. Since Eqs (15) and (17) provide equivalent estimates, as measured by mean absolute difference and largest absolute difference, the less complicated of the two equations, Eq (15), was chosen as the best plotting position with which to plot the DMAX values.

At this point, it has been determined that the linear estimation technique is a "good" estimation technique; how-

ever, substituting $S_n(x)$ in calculating the DMAX values with some other plotting positions may improve the estimates.

In determining if it would be wise to replace $S_n(x)$ with some other plotting position in calculating the DMAX values, the plotting positions defined by Eqs (15), (16), and (17) were selected as alternatives. Thus, at N = 750, three sets of critical values were formed by calculating DMAX values with the three alternative plotting positions and plotting all DMAX values with Eq (15). The set formed by using Eq (15) in the calculation of DMAX values produced a mean absolute difference of .0114 and a largest absolute difference of .0284. The set created by using Eq (16) had a mean absolute difference of .0102 and a largest absolute difference of .0256. Eq (17) formed a set of critical values with a mean absolute difference of .0032 and a largest absolute difference of .0080. These values are definitely not better or as good as the absolute differences obtained when $S_n(x)$ is used to calculate the DMAX values.

These results indicate that Eq (15) used to plot the DMAX values and $S_n(x)$ used to calculate the DMAX values provide the most effective linear estimation technique possible from the plotting positions investigated.

Using Eq (15) and $S_n(x)$ as indicated in the previous paragraph, a table of estimated critical values was produced at N = 750. This table can be examined in Appendix A.

The Power Study Results

In Appendix E can be found tables indicating the power of the most effective linear estimation technique at N = 100, 250, 500, and 750. Also, in the same tables are the power values of the standard technique. This allows a direct comparison between the powers of the standard and linear estimation techniques.

The power values for the standard technique indicate that the K-S test offers, among the distributions examined, the best power against the chi-square distribution (1 degree of freedom) and the worst power against the chi-square distribution (4 degrees of freedom). As might be expected, the power against all the distributions is poor at n = 10.

The power values given by the linear estimation technique approach those of the standard technique as N gets larger. In order to view the effect on the power of the linear estimation technique as N increases, plots of power versus N are presented in Appendix F. In these plots, the dashed horizontal lines and the full horizontal

lines depict the power associated with the significance levels .01 and .05, respectively.

Examining the plots in Appendix F, it can be seen that each set of plots related to sample size n has basically the same form. This indicates that the manner in which the power varies as N increases does not depend on the true underlying distribution of the sample but on the sample size n. It is n which determines if the critical value estimated by the linear estimation technique at some N will be smaller or larger than the standard critical value. If the critical value is estimated to be smaller (larger) than the standard critical value, the significance level will be higher (lower) than the standard and the power will also be higher (lower) than the standard.

Significance Levels

Appendix C presents the proportion of rejections when the hypothesized distribution, the extreme value distribution (largest extreme value), is the actual underlying distribution. In other words, Monte Carlo generated significance levels are tabled. It can be seen that the standard technique does well in meeting the claimed significance levels. The linear estimation technique at N = 100 does poorly; however, at N = 750, it does well.

In Appendix D are plots of the significance level values versus N. The dashed and full lines indicate the significance level values for .01 and .05, respectively. These plots present the effect increases in N have on the significance level values. Generally, these plots indicate that as N increases the significance level values approach the standard significance levels.

Chapter VI Summary

This chapter began by showing how the major computer program used in this work was validated as to giving correct critical values. Then, the results of the study were presented. These results and the conclusions stemming from them will be presented in the next and final chapter.

VII. Conclusions and Recommendations

The results introduced in the preceeding chapter leads to the conclusions offered here. In addition to the conclusions, recommendations for extensions of this thesis are given.

Conclusions

The following conclusions were drawn from the results:

1. The standard set of critical values that were generated are valid.

2. The most effective linear estimation technique is the one which Eq (15) is used to plot the DMAX values, $S_n(x)$ is used in calculating DMAX, and N = 750.

3. The power or significance levels are not altered extensively by the most effective linear estimation technique.

These conclusions can be written in terms of the first two objectives set out in Chapter 1.

1. The linear estimation technique <u>does</u> reduce the DMAX sample size required to find suitable critical values.

2. Eq (15) is better than $S_n(x)$ in plotting, the DMAX values; however, no other plotting position, of those studied, were better than $S_n(x)$ in calculating DMAX.

And, of course, the third and final objective in Chapter 1 was met by producing a table of estimated critical values using the most effective linear estimation technique.

Recommendations

Several recommendations for expanding this study are proposed.

1. An investigation could be made of the value of different interpolation techniques to form critical value estimates.

2. Plotting positions not presented in this thesis could be examined for their usefulness in an estimating technique.

3. A search could be made to find a functional relationship between N and the change in power or significance level.

4. An examination of the linear estimating technique could be made when all the parameters of the hypothesized distribution are known.

Bibliography

1. Blom, G. Statistical Estimates and Transformed Beta-Variables. New York: John Wiley and Sons. Inc., 1958.

2. Birnbaum, Z.W. "Numerical Tabulation of the Distribution for Finite Sample Size," Journal of the American Statistical Association, 47: 425-441 (1952).

3. Bogdanoff. "Earthquake Effects in the Safety and Reliability Analysis of Engineering Structures," International Conference on Structural Safety and Reliability, New York: Pergamon Press, 1972.

4. Branger. "Life Estimation and Prediction of Fighter Aircraft," International Conference on Structural Safety and Reliability, New York: Pergamon Press, 1972.

5. Chandra, M., N.D. Singpurwalla, and M.A. Stephens. Goodness of Fit Tests for the Weibull and the Extreme Value Distribution with Estimated Parameters. Washington, D.C.: Institute for Management Science and Engineering, The George Washington University, 1979.

6. Conover, W.J. Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc., 1980.

7. Cortes, R. "A Modified Kolmogorov-Smirnov Test for the Gamma and Weibull Distribution with Unknown Location and Scale Parameters." Unpublished MS thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, December 1980.

8. David, F.N. and N.L. Johnson. "The Probability Integral Transformation when Parameters are Estimated from the Sample," Biometrika, 35: 182-190 (1948).

9. Eikman, K.E. "Unbiased nearly best Linear Estimates of the Scale and Location Parameters of the Extreme Value Distribution by the use of Order Statistics." Unpublished MS thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, December 1968.

10. Gumbel, E.J. Chapter 12c.3 ("Statistical Estimation of the Endurance Limit"), Contributions to Order Statistics, New York: John Wiley and Sons, Inc., 1962.

11. ----. "The Return Period of Flood Flows," Annals of Mathematical Statistics, 12: 163-190 (1941).

12. Gumbel, E.J. and C.K. Mustafi. Comments to: Edward C. Posner, "The Application of Extreme Value Theory to Error Free Communication," Technometrics, 8: 363-366 (1966).

13. Johnson, N.L. and S. Kotz. Continuous Univariate Distributions - I. Boston: Houghton Mifflin Company, 1970.

14. Johnson, J.W. "A Modified Double Monte Carlo Technique to Approximate Reliability Confidence Limits of Systems with Components Characterized by the Weibull Distribution." Unpublished MS thesis. Air Force Institue of Technology, Wright-Patterson Air Force Base, Ohio, December 1980.

15. Kapur, K.C. and L.R. Lamberson. Reliability in Engineering Design. New York: John Wiley and Sons, Inc., 1977.

16. Kimball, B.F. "On the Choice of Plotting Positions on Probability Paper," Journal of the American Statistical Association, 55: 546-560 (1960).

17. Lilliefors, H. "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, 64: 387-389 (1969).

18. ----. "On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown," Journal of the American Statistical Association, 64: 387-389 (1969).

19. Littell, R.C., J.T. McClave, and W.W. Offen. "Goodness-of-Fit Tests for the Two Parameter Weibull Distribution," Commun. Statist. - Simula. Computa., B8(3): 257-269 (1979).

20. Massey, F.J. "The Kolmogorov-Smirnov Test for Goodness of Fit," Journal of the American Statistical Association, 46: 68-78 (1951).

21. Mendenhall, W. and R.L. Scheaffer. Mathematical Statistics with Applications. Belomnt, Calif.: Wadsworth Publishing Company, Inc., 1973.

22. Shellnut, J.W. "Conditional Linear Estimation of the Scale Parameter of the Extreme Value Distribution by the use of Selected Order Statistics." Unpublished MS thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, December 1966.

23. Sarhan, G.E. and B.G. Greenberg. Contributions to Order Statistics. New York: John Wiley and Sons, Inc., 1962.

24. Stephens, M.A. "Goodness of Fit for the Extreme Value Distribution," Biometrika, 64: 583-588 (1977).

Appendix A

Tables for the Standard Technique and the Most Effective Linear Estimation Technique

Table IV

Standard Critical Values

Sample Size n	ALPHA LEVEL				
	.20	.15	.10	.05	.01
10	.2204	.2304	.2441	.2632	.3026
15	.1808	.1898	.2009	.2194	.2534
20	.1591	.1659	.1756	.1913	.2213
25	.1436	.1502	.1586	.1720	.2003
30	.1302	.1367	.1454	.1573	.1820
35	.1215	.1267	.1343	.1472	.1700
40	.1141	.1190	.1253	.1363	.1610

Table V

Most Effective Linear Estimation Technique Critical Values

Sample Size n	ALPHA LEVEL				
	.20	.15	.10	.05	.01
4	.3166	.3297	.3419	.3740	.4655
6	.2699	.2835	.2986	.3234	.3714
8	.2360	.2492	.2685	.2841	.3308
10	.2203	.2316	.2439	.2607	.3051
12	.1998	.2090	.2257	.2402	.2747
14	.1846	.1921	.2037	.2202	.2508
16	.1736	.1809	.1942	.2101	.2417
18	.1641	.1713	.1814	.1972	.2398
20	.1592	.1665	.1766	.1908	.2228
25	.1412	.1497	.1584	.1704	.1990
30	.1322	.1390	.1459	.1570	.1824
35	.1200	.1251	.1354	.1465	.1733
40	.1135	.1185	.1275	.1387	.1609
45	.1065	.1109	.1177	.1274	.1447
50	.1018	.1072	.1135	.1270	.1407
55	.0963	.1027	.1088	.1172	.1342
60	.0931	.0993	.1048	.1146	.1285

Appendix B

<u>Tables for the Linear Estimation Technique where DMAX Values are Calculated with $S_n(x)$</u>

Table VI

Linear Estimation Technique N = 100 (a)

Sample Size n	SIGNIFICANCE LEVEL				
	.20	.15	.10	.05	.01
DMAX plotted with $S_n(x)$			DMAX calculated with $S_n(x)$		
10	.2220	.2354	.2487	.2576	.2988
25	.1427	.1521	.1622	.1694	.1907
40	.1103	.1144	.1275	.1321	.1445
LARGEST DIFFERENCE= .0165			MEAN DIFFERENCE= .0049		
DMAX plotted with Eq (13)			DMAX calculated with $S_n(x)$		
10	.2242	.2357	.2492	.2579	.3072
25	.1431	.1553	.1639	.1704	.1942
40	.1118	.1146	.1277	.1326	.1498
LARGEST DIFFERENCE= .0112			MEAN DIFFERENCE= .0043		
DMAX plotted with Eq (14)			DMAX calculated with $S_n(x)$		
10	.2237	.2356	.2491	.2578	.3047
25	.1430	.1545	.1634	.1701	.1932
40	.1115	.1146	.1276	.1325	.1482
LARGEST DIFFERENCE= .0128			MEAN DIFFERENCE= .0043		

Table VII
Linear Estimation Technique N = 100 (b)

Sample Size n	SIGNIFICANCE LEVEL				
	.20	.15	.10	.05	.01
DMAX plotted with Eq (15)			DMAX calculated with $S_n(x)$		
10	.2234	.2356	.2490	.2578	.3030
25	.1429	.1540	.1631	.1699	.1925
40	.1113	.1146	.1276	.1324	.1472
LARGEST DIFFERENCE= .0138			MEAN DIFFERENCE= .0042		
DMAX plotted with Eq (16)			DMAX calculated with $S_n(x)$		
10	.2236	.2356	.2491	.2578	.3041
25	.1430	.1543	.1633	.1700	.1929
40	.1114	.1146	.1276	.1324	.1478
LARGEST DIFFERENCE= .0132			MEAN DIFFERENCE= .0043		
DMAX plotted with Eq (17)			DMAX calculated with $S_n(x)$		
10	.2234	.2356	.2490	.2577	.3030
25	.1429	.1540	.1631	.1699	.1925
40	.1112	.1146	.1276	.1323	.1471
LARGEST DIFFERENCE= .0139			MEAN DIFFERENCE= .0043		

Table VIII

Linear Estimation Technique N = 250 (a)

Sample Size n	SIGNIFICANCE LEVEL				
	.20	.15	.10	.05	.01
DMAX plotted with $S_n(x)$			DMAX calculated with $S_n(x)$		
10	.2156	.2267	.2425	.2553	.2904
25	.1340	..1460	.1548	.1671	.2015
40	.1110	.1154	.1247	.1342	.1472
LARGEST DIFFERENCE= .0138			MEAN DIFFERENCE= .0052		
DMAX plotted with Eq (13)			DMAX calculated with $S_n(x)$		
10	.2159	.2275	.2427	.2563	.3030
25	.1403	.1464	.1554	.1678	.2115
40	.1113	.1160	.1250	.1353	.1501
LARGEST DIFFERENCE= .0112			MEAN DIFFERENCE= .0040		
DMAX plotted with Eq (14)			DMAX calculated with $S_n(x)$		
10	.2158	.2271	.2426	.2560	.3005
25	.1402	.1462	.1553	.1675	.2098
40	.1113	.1157	.1250	.1347	.1500
LARGEST DIFFERENCE= .0110			MEAN DIFFERENCE= .0042		

Table IX

Linear Estimation Technique N = 250 (b)

Sample Size n	SIGNIFICANCE LEVEL				
	.20	.15	.10	.05	.01
DMAX plotted with Eq (15)			DMAX calculated with $S_n(x)$		
10	.2157	.2268	.2426	.2557	.2988
25	.1402	.1461	.1551	.1674	.2087
40	.1112	.1154	.1249	.1344	.1499
LARGEST DIFFERENCE- .0111			MEAN DIFFERENCE- .0043		
DMAX plotted with Eq (16)			DMAX calculated with $S_n(x)$		
10	.2158	.2270	.2426	.2559	.2999
25	.1402	.1462	.1552	.1675	.2094
40	.1113	.1156	.1249	.1346	.1499
LARGEST DIFFERENCE- .0111			MEAN DIFFERENCE- .0042		
DMAX plotted with Eq (17)			DMAX calculated with $S_n(x)$		
10	.2157	.2268	.2426	.2557	.2988
25	.1402	.1461	.1551	.1674	.2087
40	.1112	.1154	.1249	.1344	.1498
LARGEST DIFFERENCE- .0112			MEAN DIFFERENCE- .0043		

Table X

Linear Estimation Technique N = 500 (a)

Sample Size n	SIGNIFICANCE LEVEL				
	.20	.15	.10	.05	.01
DMAX plotted with $S_n(x)$			DMAX calculated with $S_n(x)$		
10	.2191	.2290	.2425	.2576	.3012
25	.1413	.1483	.1567	.1674	.2002
40	.1119	.1170	.1247	.1368	.1546
LARGEST DIFFERENCE= .0064			MEAN DIFFERENCE= .0024		
DMAX plotted with Eq (13)			DMAX calculated with $S_n(x)$		
10	.2193	.2291	.2427	.2579	.3072
25	.1415	.1487	.1568	.1683	.2064
40	.1121	.1176	.1250	.1386	.1552
LARGEST DIFFERENCE= .0061			MEAN DIFFERENCE= .0027		
DMAX plotted with Eq (14)			DMAX calculated with $S_n(x)$		
10	.2192	.2291	.2426	.2578	.3054
25	.1415	.1486	.1567	.1680	.2046
40	.1121	.1174	.1250	.1381	.1550
LARGEST DIFFERENCE= .0060			MEAN DIFFERENCE= .0025		

Table XI

Linear Estimation Technique N = 500 (b)

Sample Size n	SIGNIFICANCE LEVEL .20	.15	.10	.05	.01
DMAX plotted with Eq (15)			DMAX calculated with $S_n(x)$		
10	.2192	.2291	.2426	.2578	.3042
25	.1414	.1485	.1567	.1679	.2033
40	.1120	.1173	.1249	.1377	.1549
LARGEST DIFFERENCE= .0061			MEAN DIFFERENCE= .0024		
DMAX plotted with Eq (16)			DMAX calculated with $S_n(x)$		
10	.2192	.2291	.2426	.2578	.3050
25	.1414	.1486	.1567	.1680	.2041
40	.1121	.1174	.1249	.1379	.1550
LARGEST DIFFERENCE= .0060			MEAN DIFFERENCE= .0025		
DMAX plotted with Eq (17)			DMAX calculated with $S_n(x)$		
10	.2192	.2291	.2426	.2578	.3042
25	.1414	.1485	.1567	.1679	.2033
40	.1120	.1173	.1249	.1377	.1549
LARGEST DIFFERENCE= .0061			MEAN DIFFERENCE= .0024		

Table XII

Linear Estimation Technique N = 750 (a)

Sample Size n	SIGNIFICANCE LEVEL .20	.15	.10	.05	.01
DMAX plotted with $S_n(x)$			DMAX calculated with $S_n(x)$		
10	.2202	.2315	.2437	.2605	.3031
25	.1412	.1496	.1583	.1700	.1979
40	.1134	.1184	.1275	.1382	.1590
LARGEST DIFFERENCE= .0027			MEAN DIFFERENCE= .0013		
DMAX plotted with Eq (13)			DMAX calculated with $S_n(x)$		
10	.2203	.2317	.2440	.2608	.3061
25	.1413	.1498	.1584	.1706	.1996
40	.1135	.1185	.1276	.1389	.1626
LARGEST DIFFERENCE= .0035			MEAN DIFFERENCE= .0013		
DMAX plotted with Eq (14)			DMAX calculated with $S_n(x)$		
10	.2203	.2316	.2439	.2608	.3055
25	.1412	.1498	.1584	.1705	.1992
40	.1135	.1185	.1275	.1388	.1616
LARGEST DIFFERENCE= .0029			MEAN DIFFERENCE= .0013		

Table XIII

Linear Estimation Technique N = 750 (b)

Sample Size n	SIGNIFICANCE LEVEL				
	.20	.15	.10	.05	.01
DMAX plotted with Eq (15)			DMAX calculated with $S_n(x)$		
10	.2203	.2316	.2439	.2607	.3051
25	.1412	.1497	.1584	.1704	.1990
40	.1135	.1185	.1275	.1387	.1609
LARGEST DIFFERENCE= .0025			MEAN DIFFERENCE= .0012		
DMAX plotted with Eq (16)			DMAX calculated with $S_n(x)$		
10	.2203	.2316	.2439	.2607	.3053
25	.1412	.1497	.1584	.1705	.1991
40	.1135	.1185	.1275	.1387	.1613
LARGEST DIFFERENCE= .0027			MEAN DIFFERENCE= .0012		
DMAX plotted with Eq (17)			DMAX calculated with $S_n(x)$		
10	.2203	.2316	.2439	.2607	.3050
25	.1412	.1497	.1584	.1704	.1990
40	.1135	.1185	.1275	.1387	.1609
LARGEST DIFFERENCE= .0025			MEAN DIFFERENCE= .0012		

Appendix C

Table of Significance Levels

Table XIV

Significance Levels

	Standard		Extreme Value Dist. (largest extr. val.)							
			N=750		N=500		N=250		N=100	
n \ α	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01
10	.0500	.0104	.0552	.0088	.0624	.0100	.0704	.0128	.0624	.0100
25	.0504	.0100	.0552	.0116	.0620	.0092	.0628	.0060	.0568	.0164
40	.0540	.0112	.0456	.0120	.0488	.0160	.0592	.0204	.0696	.0248

Appendix D

Plot of Significance Levels

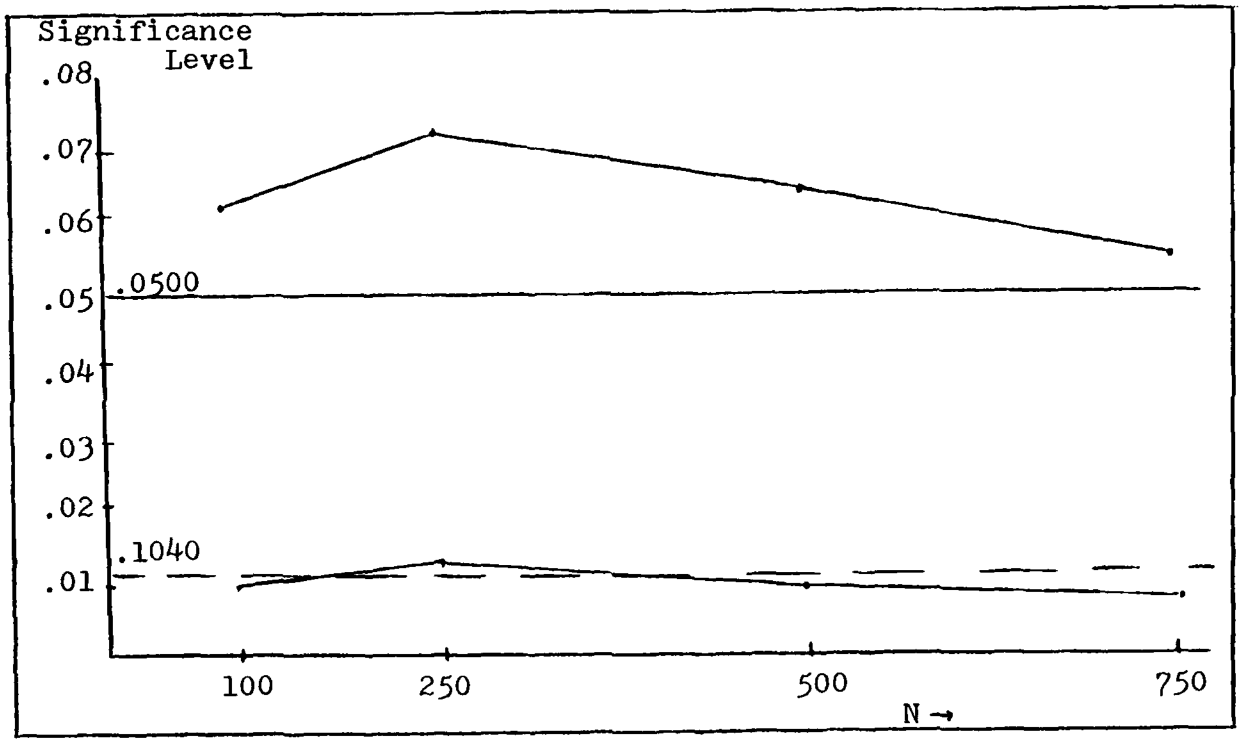

Fig 5.

Extreme Value (Largest Extreme Value) n = 10

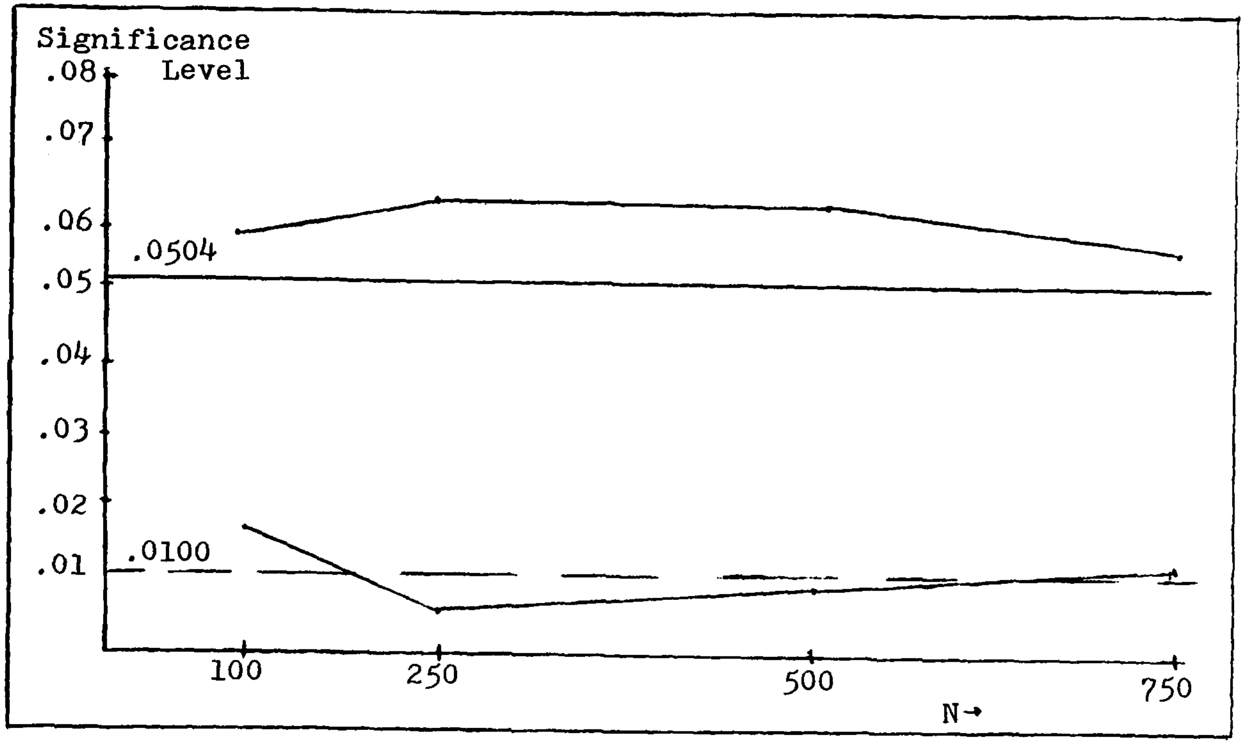

Fig 6.

Extreme Value (Largest Extreme Value) n = 25

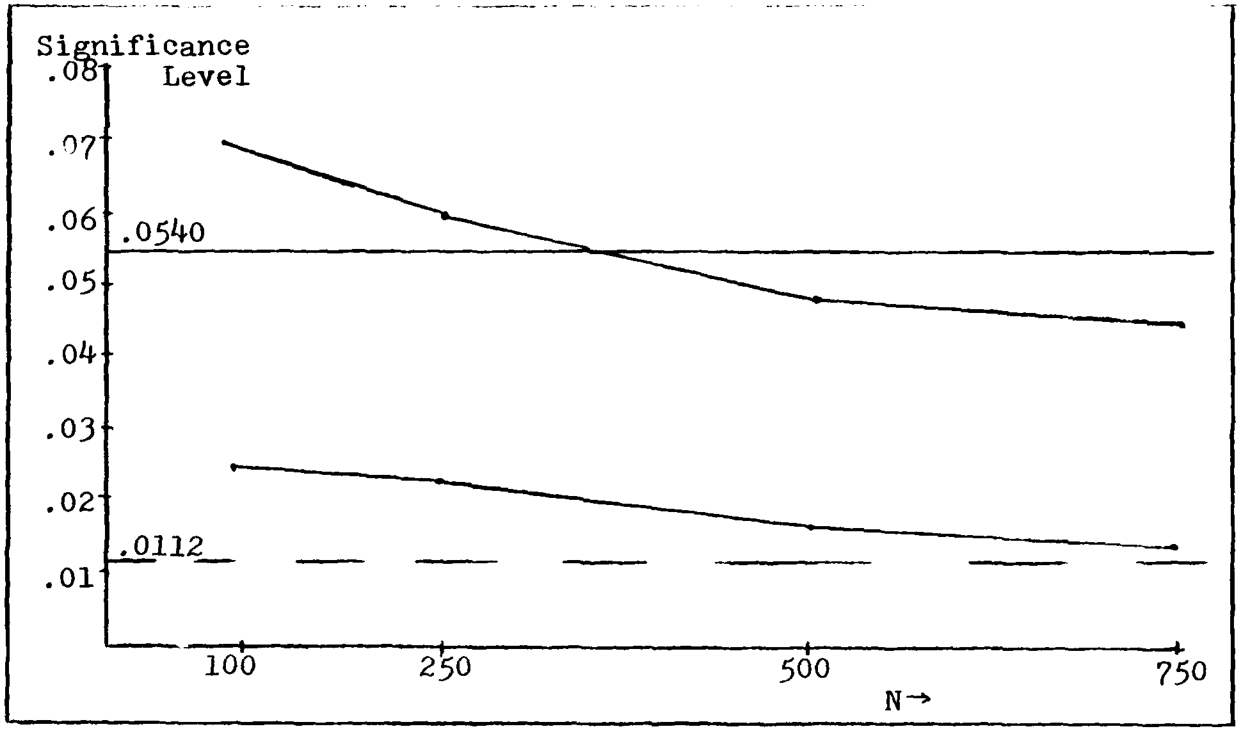

Fig 7.

Extreme Value (Largest Extreme Value) n = 40

Appendix E

Tables of Power Values

Table XV

Power Table

	Standard		Standard Normal							
			N=750		N=500		N=250		N=100	
n \ α	.05	: .01	.05	: .01	.05	: .01	.05	: .01	.05	: .01
10	.0932	.0284	.0972	.0256	.1064	.0272	.1124	.0324	.1064	.0280
25	.1916	.0716	.2012	.0748	.2204	.0628	.2228	.0492	.2044	.0984
40	.3200	.1168	.2936	.1176	.3032	.1556	.3400	.1916	.3640	.2160

Table XVI

Power Table

	Standard		Two-parameter Weibull							
			N=750		N=500		N=250		N=100	
n \ α	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01
10	.1428	.0464	.1504	.0424	.1644	.0440	.1692	.0544	.1644	.0460
25	.3008	.1260	.3144	.1316	.3380	.1132	.3412	.0928	.3212	.1608
40	.4700	.2240	.4404	.2248	.4512	.2752	.4952	.3252	.5192	.3576

Table XVII

Power Table

	Standard		Two-parameter Log-Normal							
			N=750		N=500		N=250		N=100	
n \ α	.05 :	.01	.05 :	.01	.05 :	.01	.05 :	.01	.05 :	.01
10	.2384	.1076	.2452	.1020	.2568	.1036	.2680	.1212	.2568	.1064
25	.5236	.3016	.5388	.3128	.5600	.2820	.5640	.2492	.5416	.3600
40	.7368	.5000	.7128	.5008	.7224	.5632	.7548	.6092	.7724	.6372

XVIII

Power Table

	Standard		Chi-squared (1 degree of freedom)							
			N=750		N=500		N=250		N=100	
n \ α	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01
10	.3288	.1672	.3420	.1568	.3596	.1588	.3748	.1748	.3596	.1640
25	.7628	.5208	.7732	.5328	.7900	.4908	.7936	.4464	.7748	.5282
40	.9432	.7784	.9332	.7788	.9360	.8336	.9496	.8724	.9560	.8884

Table XIX

Power Table

	Standard		Chi-square (4 degrees of freedom)							
			N=750		N=500		N=250		N=100	
n \ α	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01
10	.0584	.0112	.0620	.0112	.0668	.0112	.0708	.0128	.0668	.0112
25	.0768	.0184	.0856	.0196	.0932	.0172	.0956	.0140	.0860	.0296
40	.1052	.0316	.0896	.0316	.0968	.0436	.1152	.0544	.1276	.0644

Appendix F

Plots of Power Values

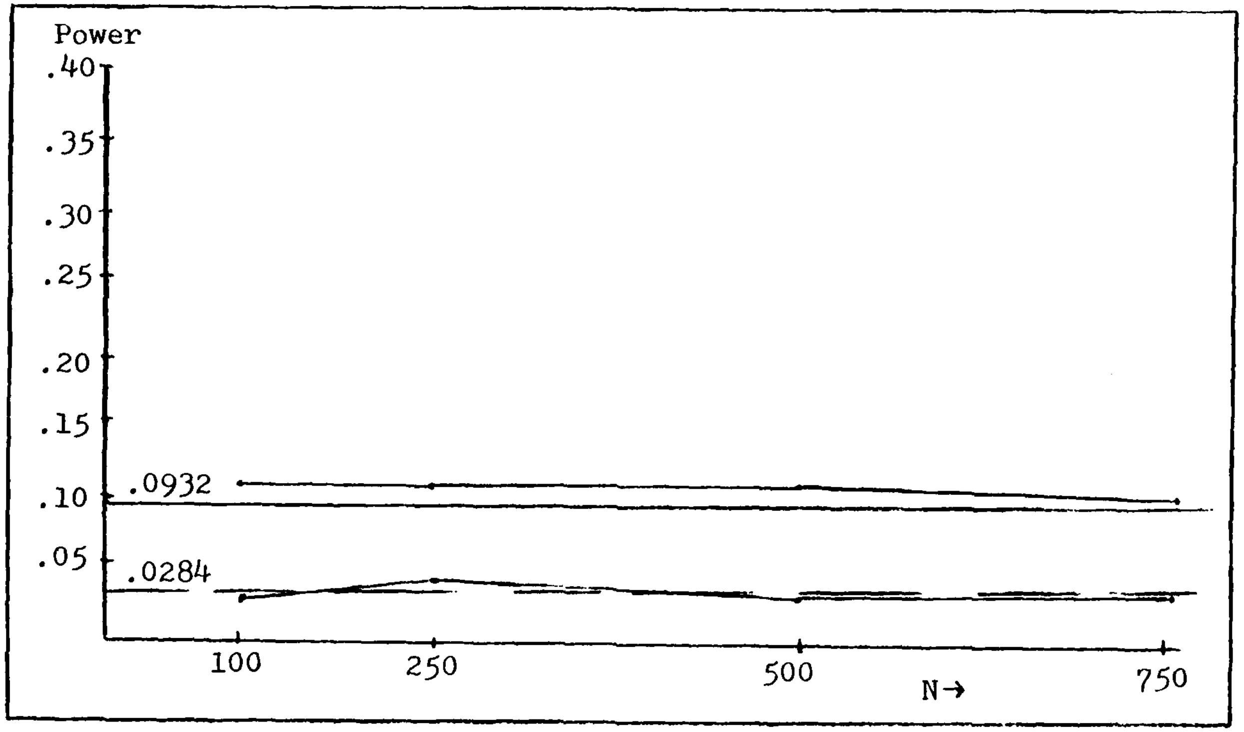

Fig 8.

Standard Normal n = 10

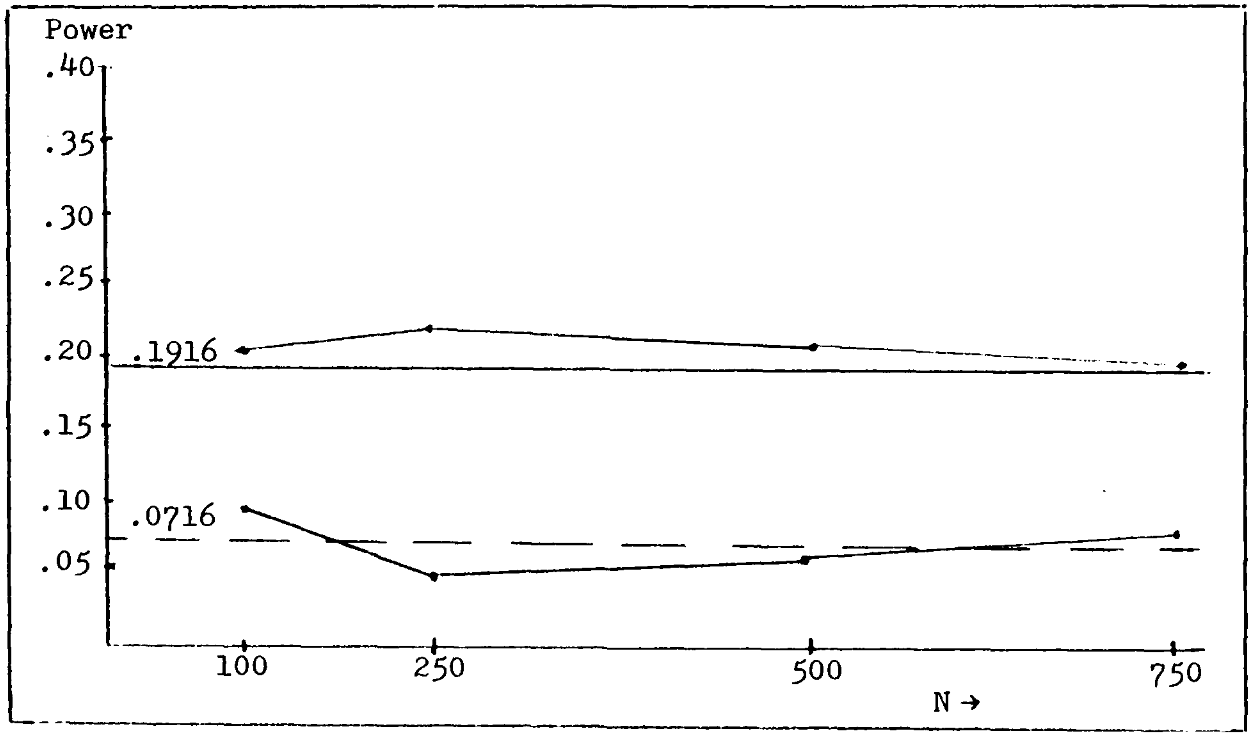

Fig 9.

Standard Normal n = 25

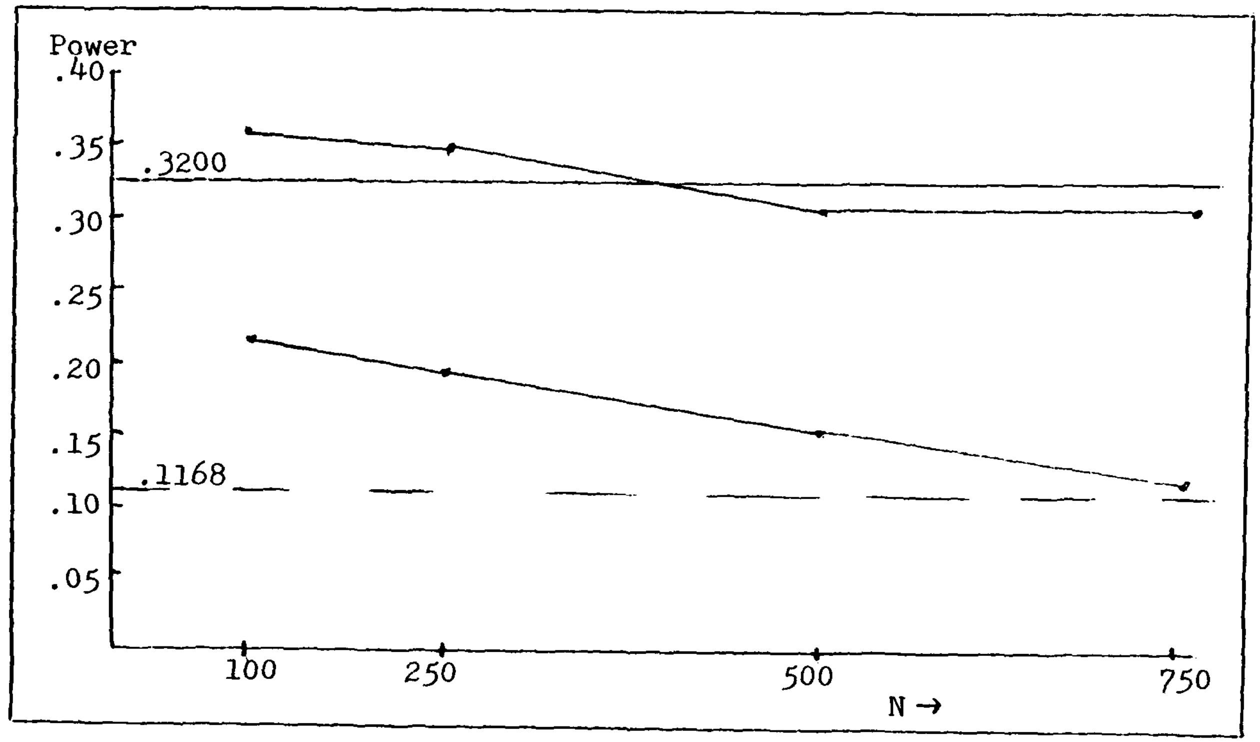

Fig 10.

Standard Normal n = 40

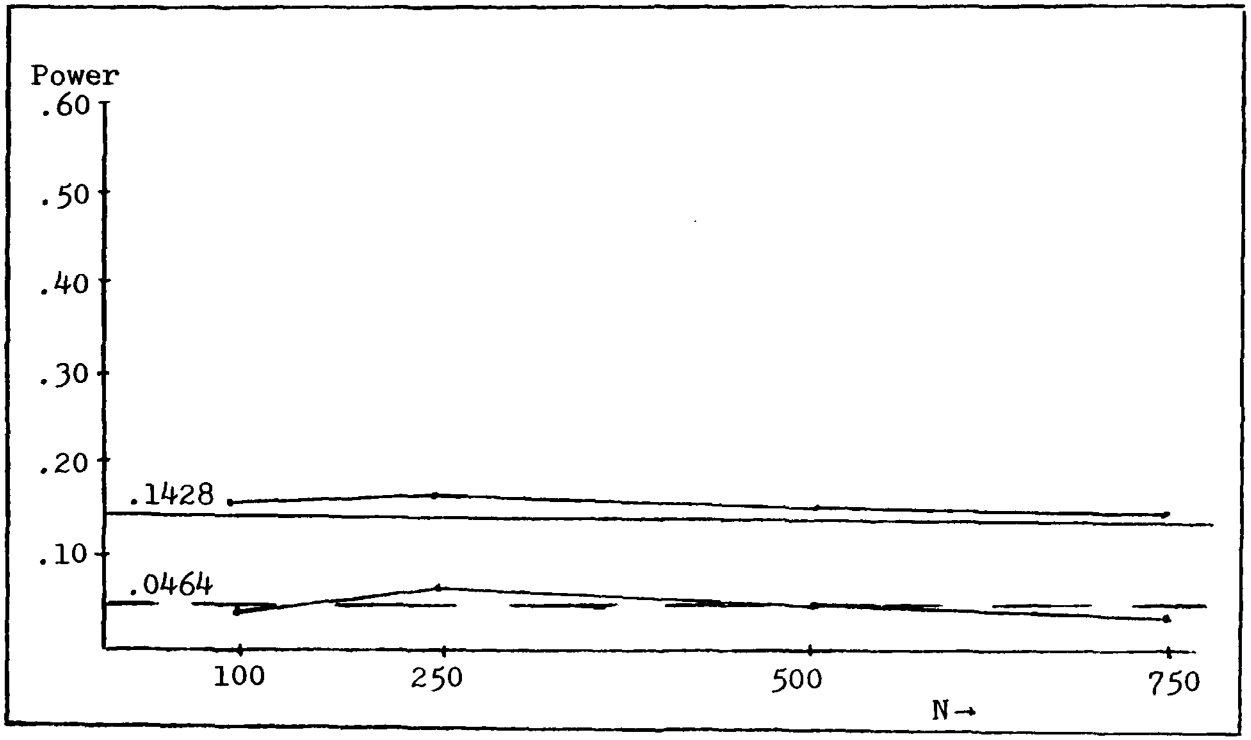

Fig 11.

Two-parameter Weibull n = 10

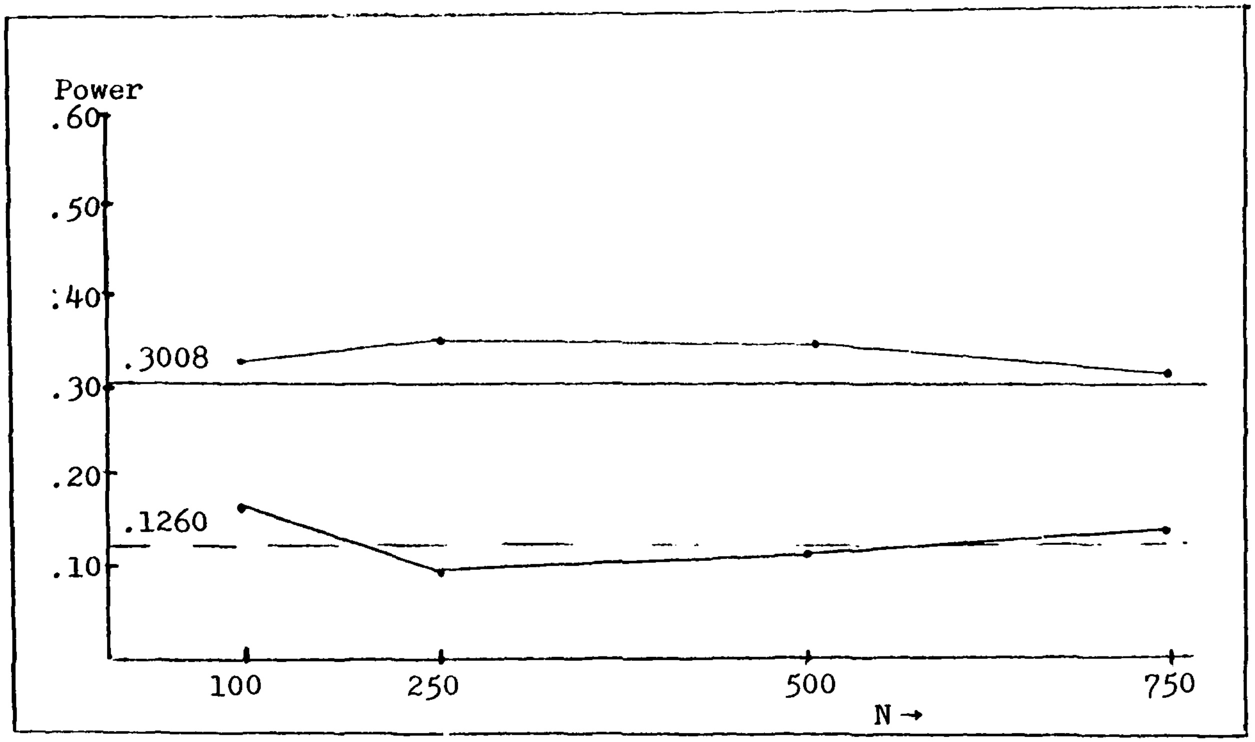

Fig 12.

Two-parameter Weibull n = 25

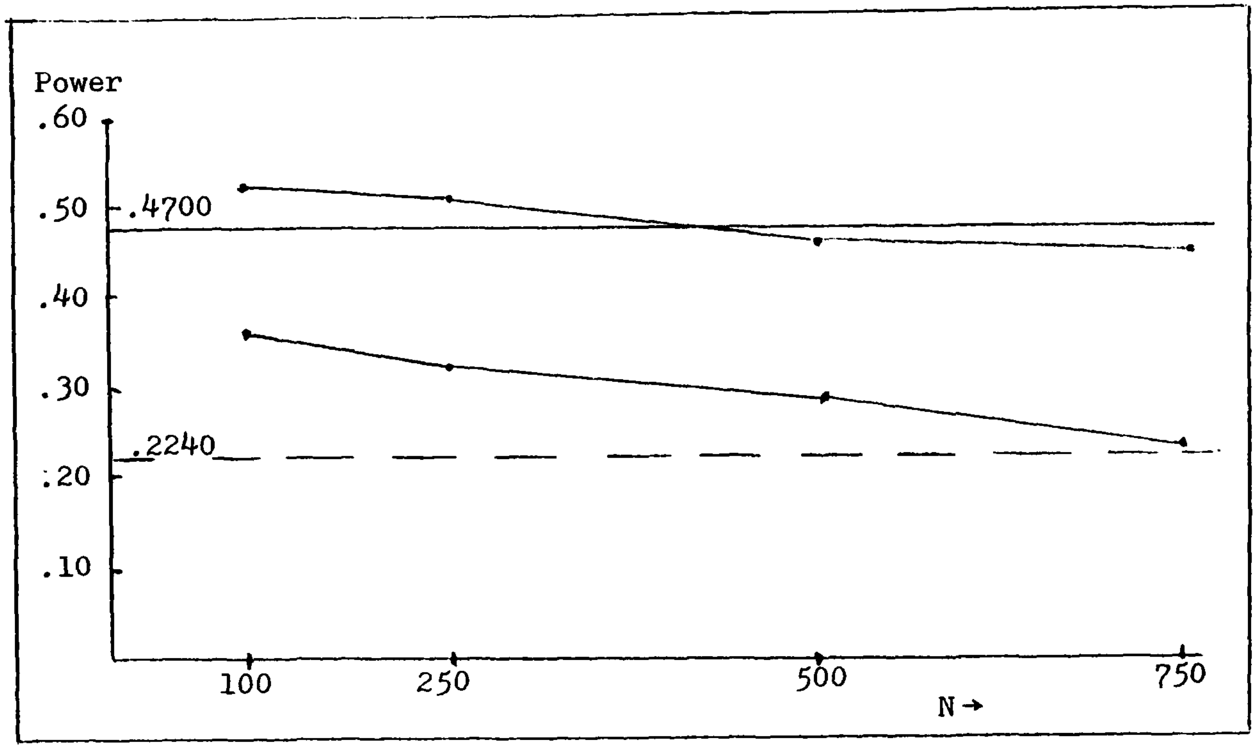

Fig 13.

Two-parameter Weibull n = 40

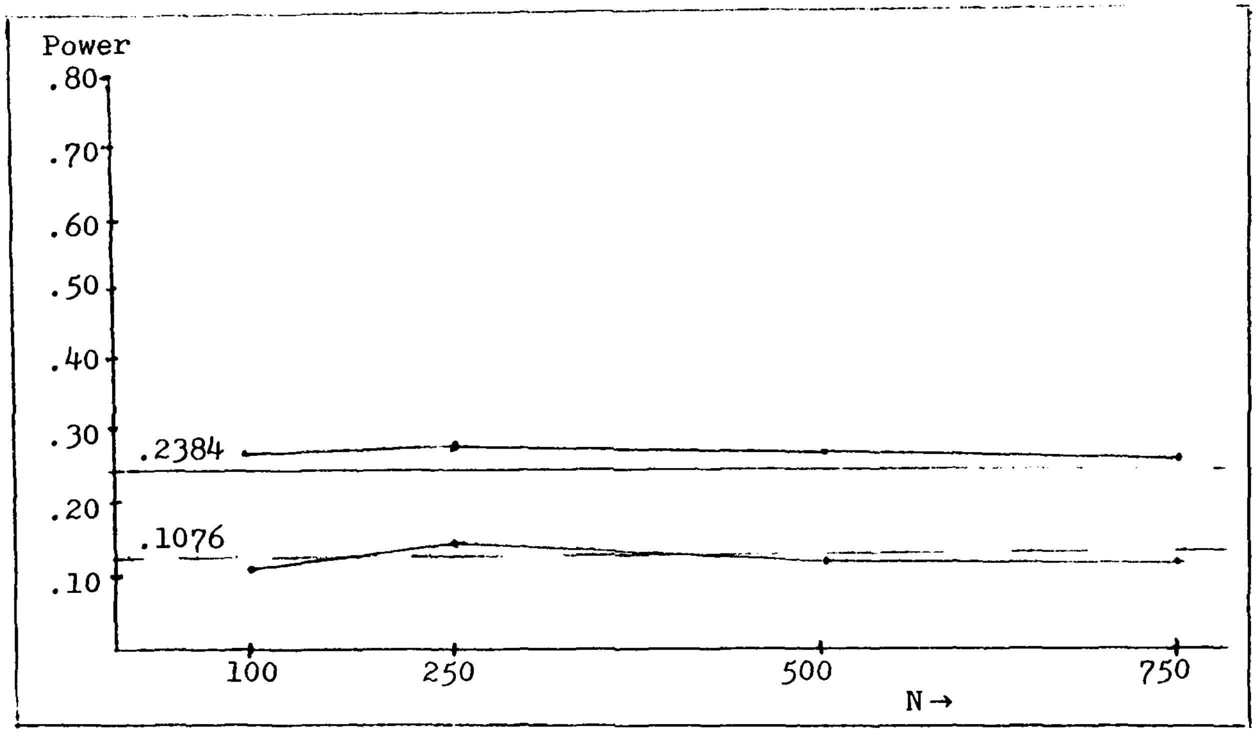

Fig 14.

Two-parameter Log-Normal n = 10

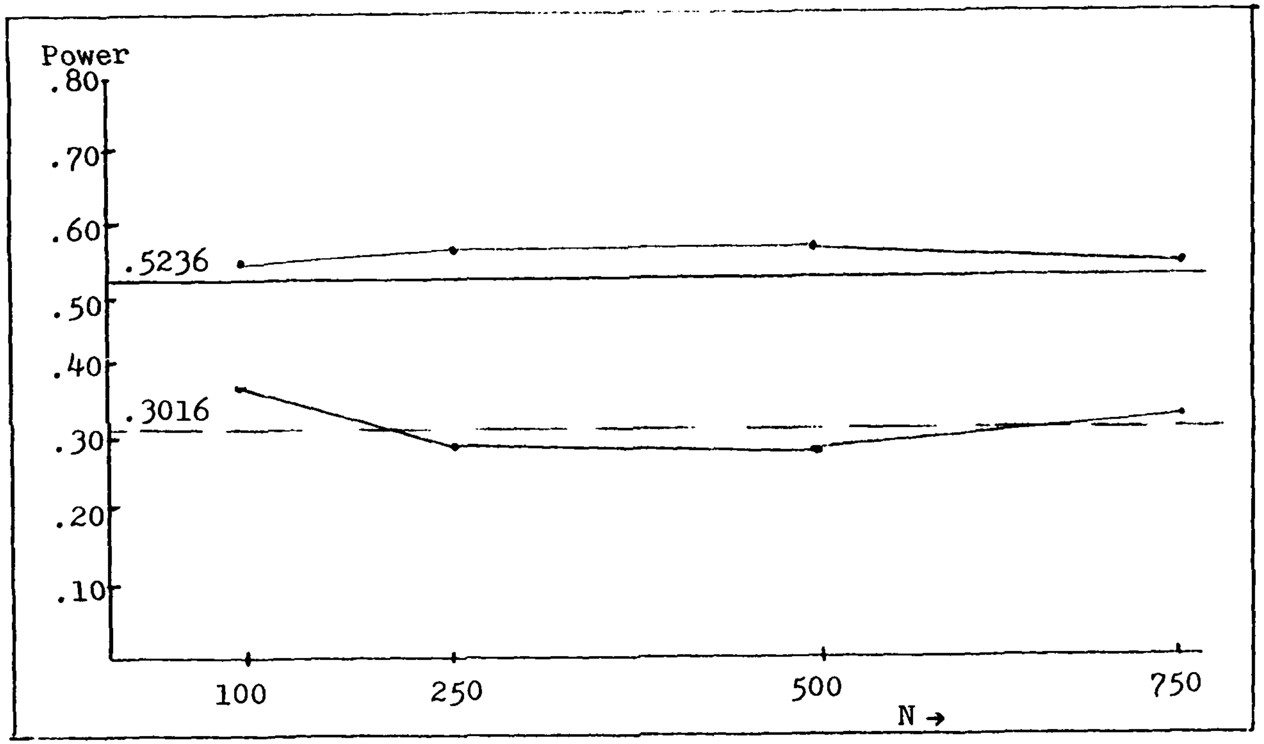

Fig 15.

Two-parameter Log-Normal n = 25

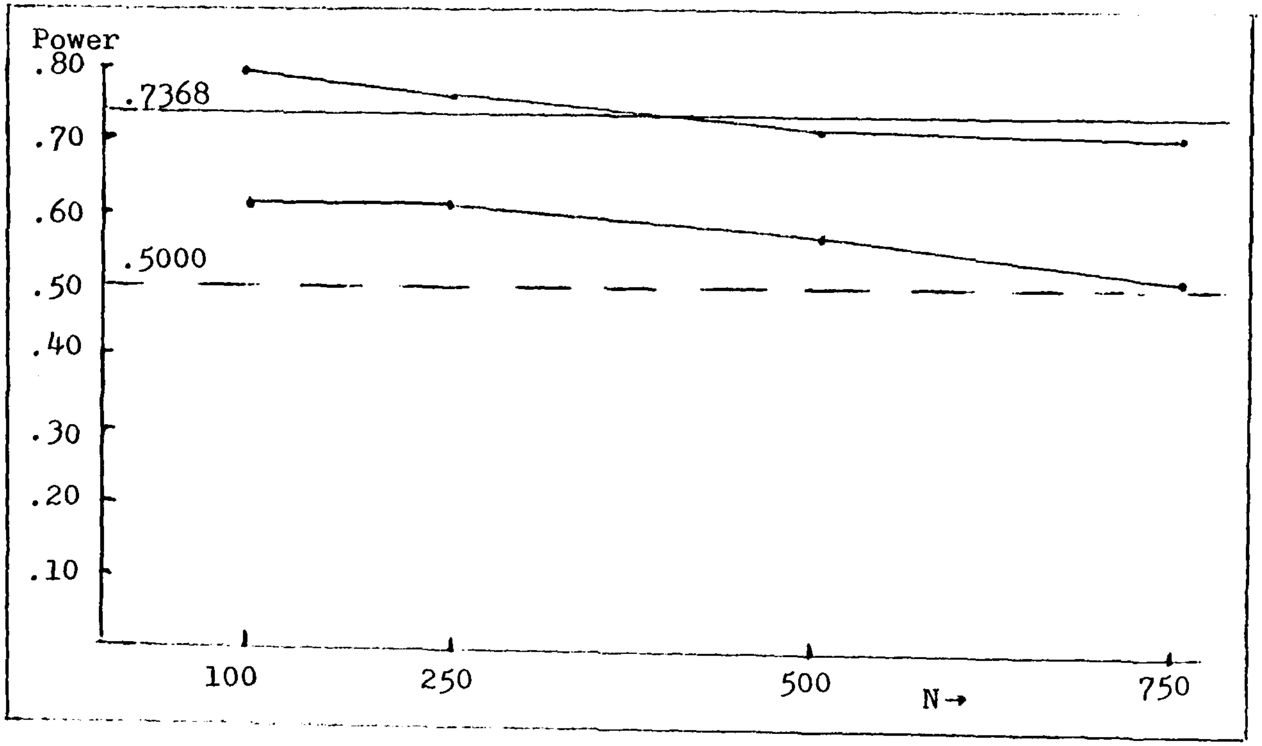

Fig 16.

Two-parameter Log-Normal $n = 40$

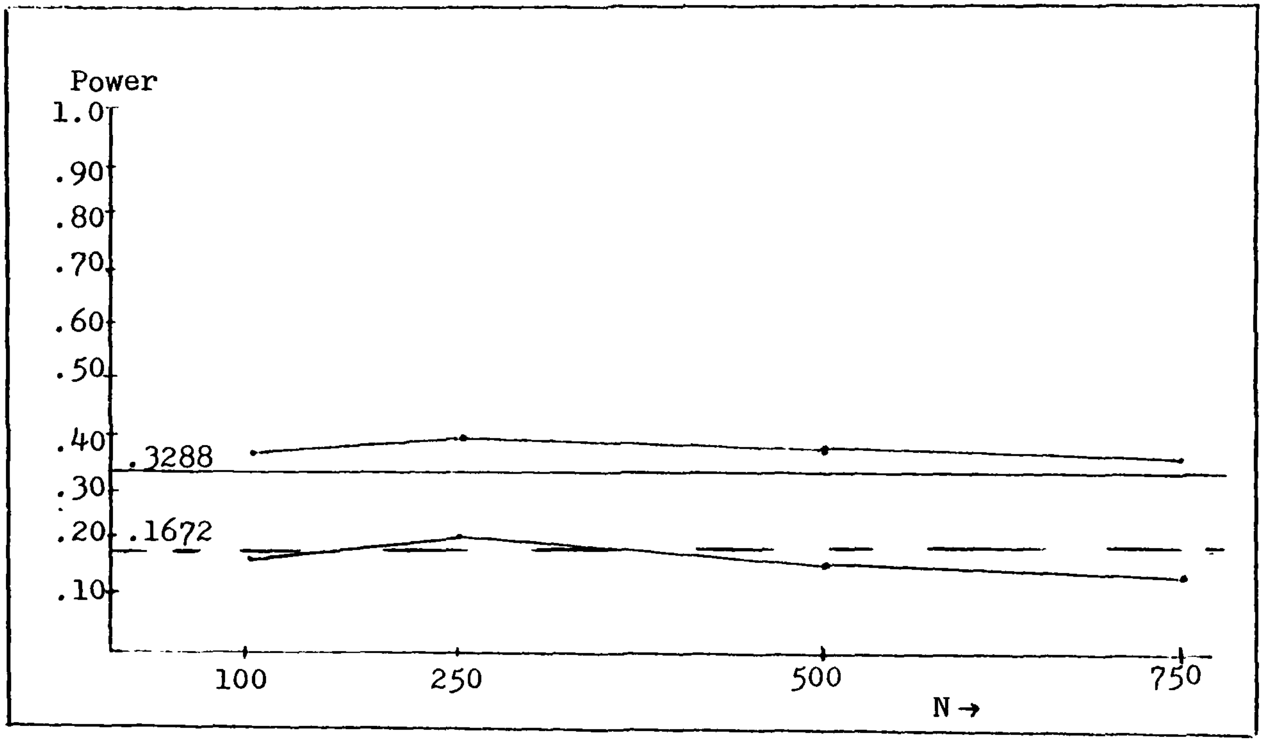

Fig 17.

Chi-square (1Degree of Freedom) n = 10

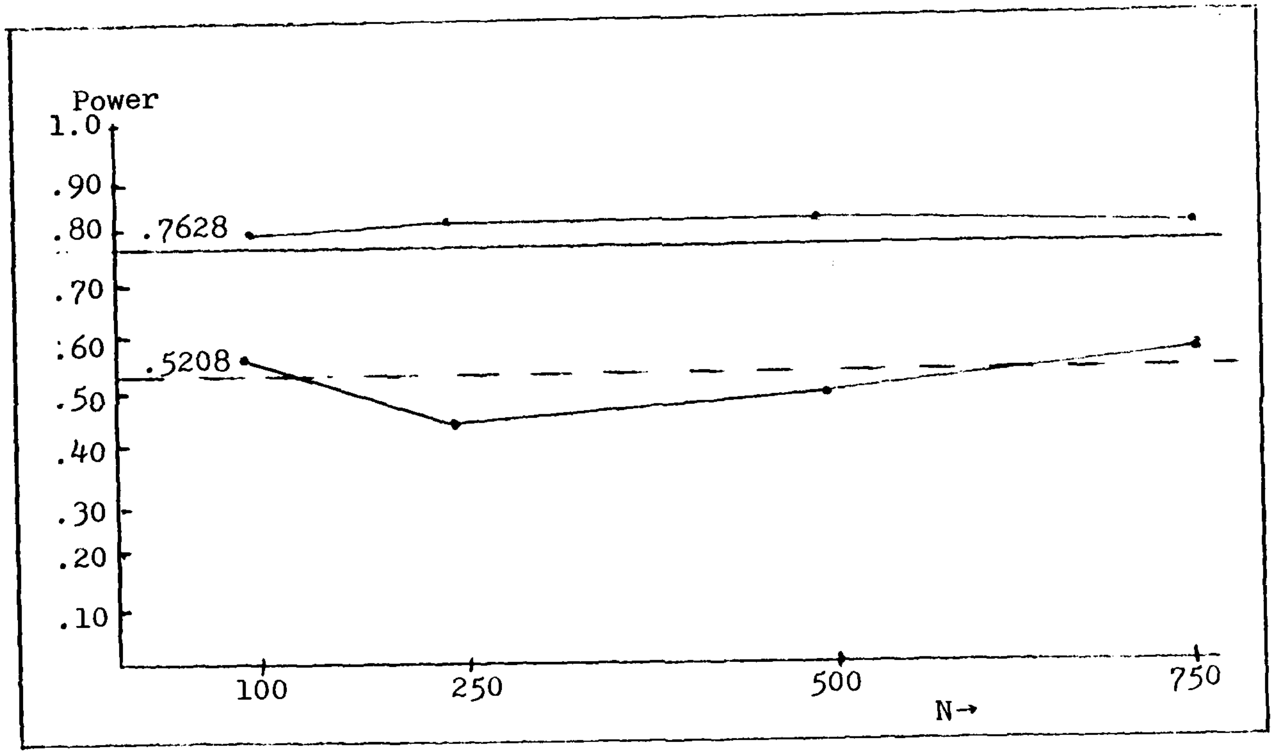

Fig 18.

Chi-square (1 Degree of Freedom) n = 25

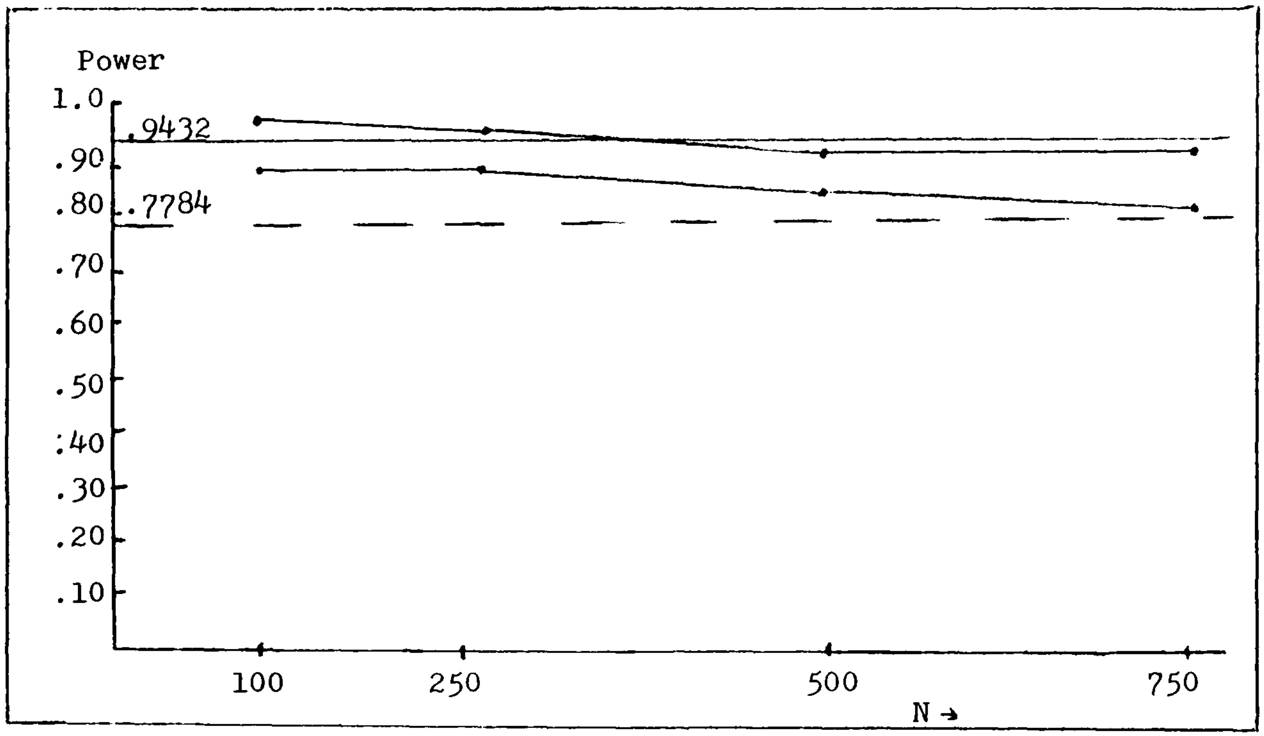

Fig 19.

Chi-square (1 Degree of Freedom) n = 40

AD-A115 544 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 12/1
A MONTE CARLO TECHNIQUE USING LINEAR INTERPOLATION TO GENERATE --ETC(U)
DEC 81 D R ROGERS
UNCLASSIFIED AFIT/GOR/MA/81D-11 NL

2 OF 2
AD A
115544

END
DATE
FILMED
7-82
DTIC

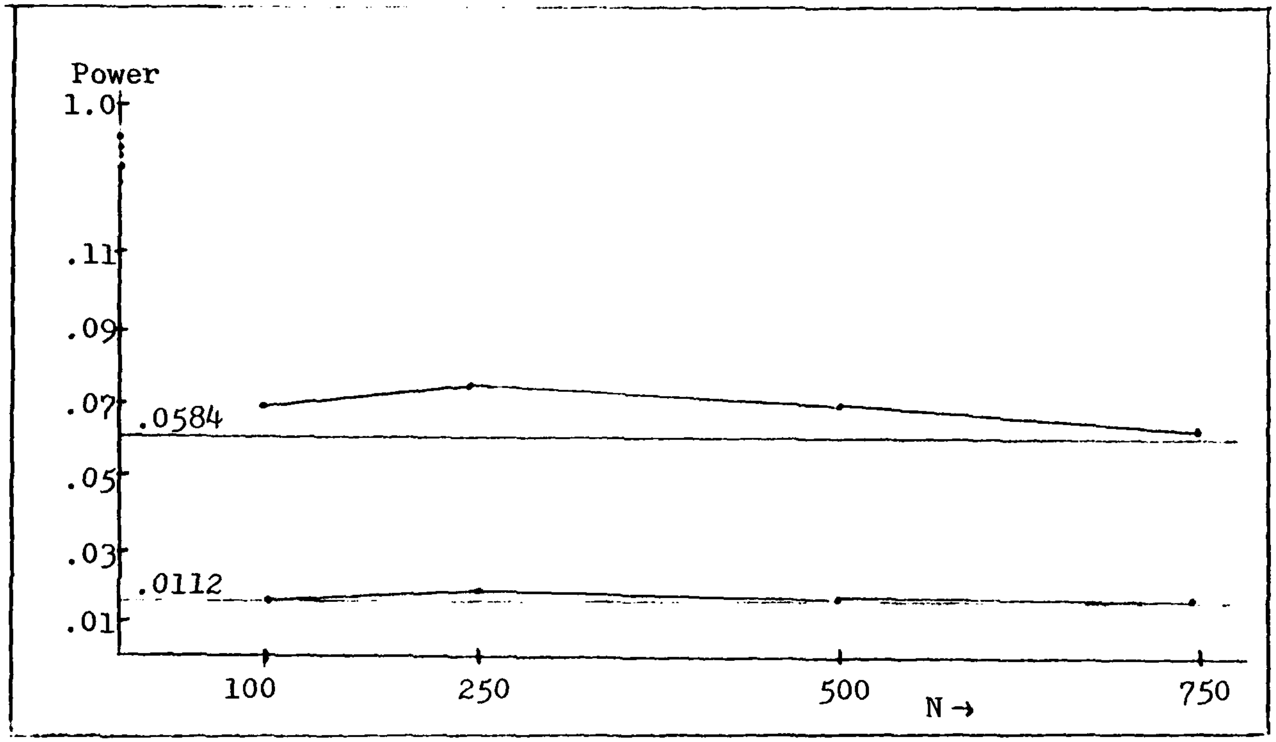

Fig 20.

Chi-square (4 Degrees of Freedom) n = 10

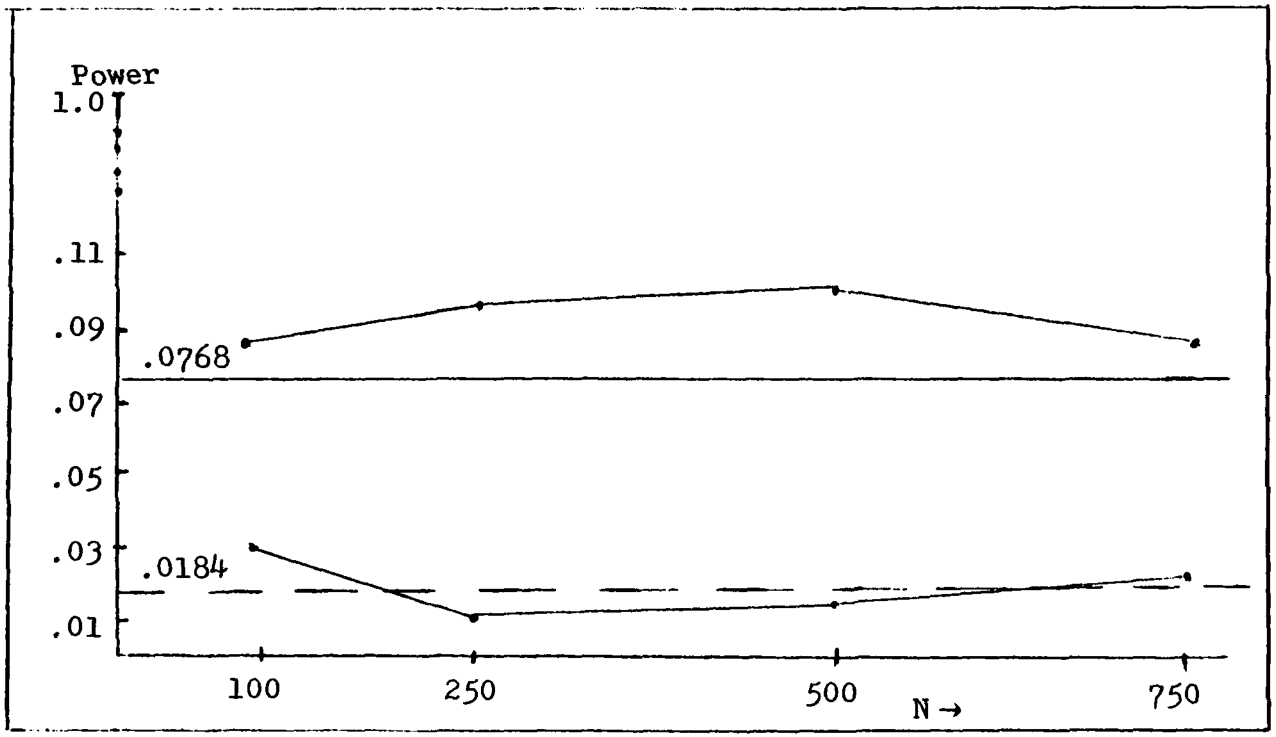

Fig 21.

Chi-square (4 Degrees of Freedom) n = 25

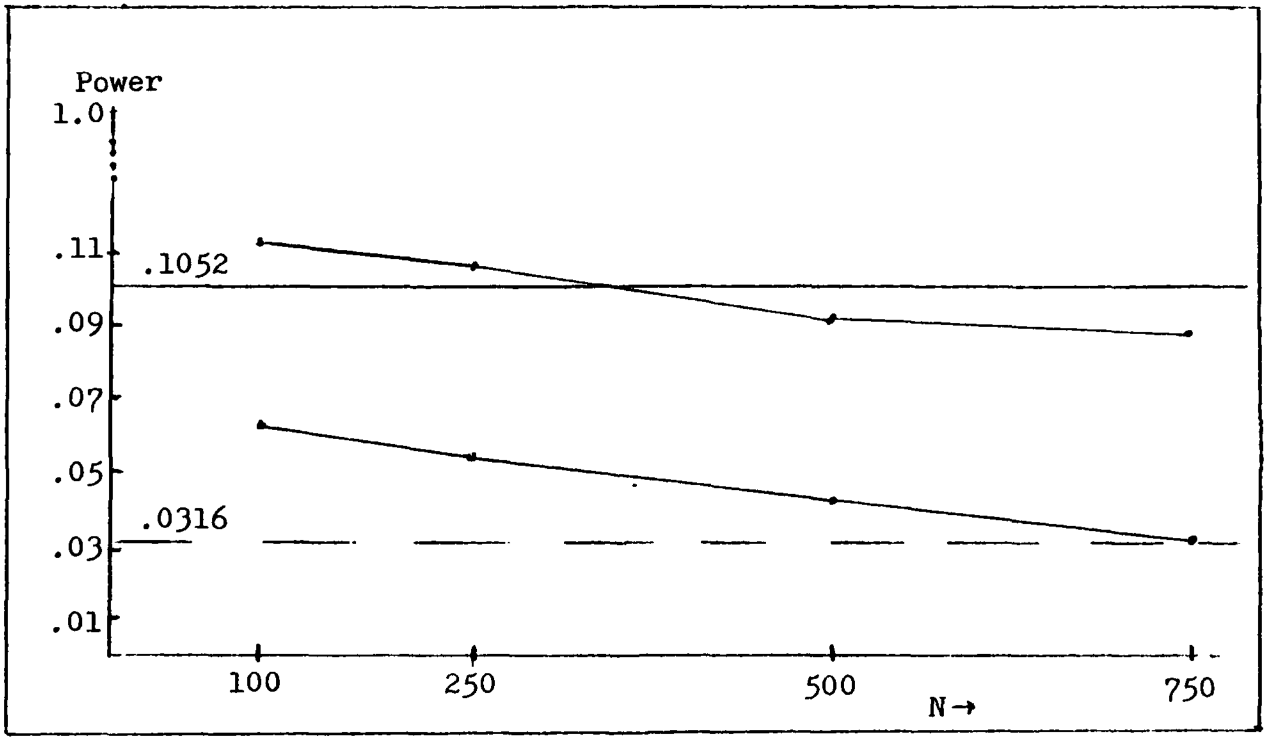

Fig 22.

Chi-Square (4 Degrees of Freedom) n = 40

Appendix G

Computer Programs

```
DRR,T500,CM65000.   T800627,ROGERS,4458
ATTACH,IMSL,ID=LIBRARY,SN=ASD.
LIBRARY,IMSL.
FTN5,ANSI=0.
LGO.
*EOR
      PROGRAM THEES2
C     **** NUMBERING OF PLOTTING POSITIONS(PLP) *******
C     * I=ITH ORDERD RANDOM DEVIATE                   *
C     * N=NUMBER OF RANDOM DEVIATES(OR DNA STATISTICS)*
C     * #1:(I/N)                                      *
C     * #2:(I-.5)/(N)                                 *
C     * #3:(I-.3)/(N+.4)                              *
C     * #4:((I/(N+1))+((I-1)/(N-1)))/2                *
C     * #5:(I-.375)/(N+.25)                           *
C     * #6:(I/(N+1))                                  *
C     *************************************************
C
C     **** CONDITION CODES *************************************
C     * COND2:PLOT D STATISTICS USING PLP#2                    *
C     * COND3:PLOT D STATISTICS USING PLP#3                    *
C     * COND4:PLOT D STATISTICS USING PLP#4                    *
C     * COND5:PLOT D STATISTICS USING PLP#5                    *
C     * COND6:PLOT D STATISTICS USING PLP#6                    *
C     * IF A CONDITON IS GIVEN ANY INTEGER VALUE OTHER THAN 1,*
C     * THE CORRESPONDING COMPUTATION WILL NOT OCCUR           *
C     *     FURTHER, SNC MUST BE ASSIGNED AN INTEGER VALUE     *
C     * FROM 1 TO 5. THE VALUE CORRESPONDS TO A PLP USED TO    *
C     * COMPUTE SN.                                            *
C     **********************************************************
C
      DIMENSION R(10),ABDIF(20),DNA(5000),SN(11)
      INTEGER REP,SNC
      DOUBLE PRECISION DSEED
      DATA A,B/2*1.0/
      C=0.0
      NRD=10
      MAXLIK=500
      REP=5000
      SNC=1
      DSEED=1234567.0D0
      DO 501 I=1,5000
C     **** GENERATES EXTR. VAL. RANDOM DEVIATES ****
      CALL GGNIB(DSEED,A,NRD,R)
      DO 1 I1=1,NRD
          R(I1)=-(LOG(B*R(I1)+C))
    1 CONTINUE
      LOGB=-(LOG(B))
      THETA=1.00/A
C     ************************************************
C     **** CALCULATES MAXIMUM LIKELIHOOD ESTIMATES ****
      SUM1=0.0
      SUM2=0.0
      SUM3=0.0
```

```
      SUM4=0.0
      DO 3 I3=1,NRD
        SUM1=R(I3)+SUM1
    3 CONTINUE
      DO 4 I4=1,MAXLIK
        DO 5 I5=1,NRD
           ZX=-(R(I5)/THETA)
           SUM2=R(I5)*EXP(ZX)+SUM2
           SUM3=EXP(ZX)+SUM3
    5   CONTINUE
      THETA1=(SUM1/NRD)-(SUM2/SUM3)
      DIF=ABS(THETA1-THETA)
      THETA=THETA1
      IF(DIF.LE.0.00001) GOTO100
    4 CONTINUE
  100 DO 6 I6=1,NRD
         ZX1=-(R(I6)/THETA)
         SUM4=EXP(ZX1)+SUM4
    6 CONTINUE
      EPS=(-THETA)*(LOG(SUM4/NRD))
C     ****************************************************
C     **** GENERATES DNA STATISTICS ****
      CALL VSRTA(R,NRD)
      SN(1)=0.0
      DO 7 I7=2,NRD+1
        XI7=REAL(I7)
        XNRD=REAL(NRD)
        IF(SNC.EQ.1) THEN
          SN(I7)=(XI7-1.0)/(XNRD)
        ENDIF
    7 CONTINUE
      DO 10 J1=1,NRD
        CDF=EXP(-EXP((EPS-R(J1))/THETA))
        ABDIF(J1)=ABS(CDF-SN(J1))
             ABDIF(J1+NRD)=ABS(CDF-SN(J1+1))
   10 CONTINUE
      NRD2=2*NRD
      CALL VSRTA(ABDIF,NRD2)
      DNA(I)=ABDIF(NRD2)
  501 CONTINUE
      CALL VSRTA(DNA,REP)
C     ************************************************
C     **** PRINTS CRITICAL VALUES FOR 5000 REPS ****
      PRINT*,'**************************************'
      PRINT*,'THE FOLLOWING CRITICAL VALUES'
      PRINT*,'WERE CALCULATED FROM 5000'
      PRINT*,'DIFFERENT SETS OF ',NRD
      PRINT*,'RANDOM DEVIATES'
      PRINT*,'**************************************'
      PRINT*,'THE 80TH%= ',DNA(4000)
      PRINT*,'THE 85TH%= ',DNA(4250)
      PRINT*,'THE 90TH%= ',DNA(4500)
      PRINT*,'THE 95TH%= ',DNA(4750)
      PRINT*,'THE 99TH%= ',DNA(4950)
      STOP
      END
```

```
DRR,T400,CM65000.   T800627,ROGERS,4458
ATTACH,IMSL,ID=LIBRARY,SN=ASD.
LIBRARY,IMSL.
FTN5,ANSI=0.
LGO.
*EOR
      PROGRAM CRIVAL
C     **** NUMBERING OF PLOTTING POSITIONS(PLP) *******
C     * I=ITH ORDERD RANDOM DEVIATE                   *
C     * N=NUMBER OF RANDOM DEVIATES(OR DNA STATISTICS)*
C     * #1:(I/N)                                      *
C     * #2:(I-.5)/(N)                                 *
C     * #3:(I-.3)/(N+.4)                              *
C     * #4:((I/(N+1))+((I-1)/(N-1)))/2                *
C     * #5:(I-.375)/(N+.25)                           *
C     * #6:(I/(N+1))                                  *
C     *************************************************
C
C     **** CONDITION CODES ***************************************
C     * COND1:PLOT D STATISTICS USING PLP#1                      *
C     * COND2:PLOT D STATISTICS USING PLP#2                      *
C     * COND3:PLOT D STATISTICS USING PLP#3                      *
C     * COND4:PLOT D STATISTICS USING PLP#4                      *
C     * COND5:PLOT D STATISTICS USING PLP#5                      *
C     * COND6:PLOT D STATISTICS USING PLP#6                      *
C     * IF A CONDITON IS GIVEN ANY INTEGER VALUE OTHER THAN 1,*
C     * THE CORRESPONDING COMPUTATION WILL NOT OCCUR            *
C     *     FURTHER, SNC MUST BE ASSIGNED AN INTEGER VALUE      *
C     * FROM 1 TO 6. THE VALUE CORRESPONDS TO A PLP USED TO     *
C     * COMPUTE SN.                                             *
C     ************************************************************
C
      DIMENSION R(40),ABDIF(80),DNA(750),Y(752),TDNA(752),SN(41)
      INTEGER REP,DIM,COND1,COND2,COND3,COND4,COND5,COND6,SNC
      DOUBLE PRECISION DSEED
      DATA A,B/2*1.0/
      C=0.0
      NRD=40
      INRD=40
      INRD2=INRD*2
      MAXLIK=500
      REP=750
      SNC=1
      COND1=1
      COND2=1
      COND3=1
      COND4=1
      COND5=1
      COND6=1
      DSEED=1234567.0D0
      IZZ=NRD
      IF(NRD.GT.5) GOTO333
      DO 500 NRD=5,INRD
  333 DO 501 I=1,REP
```

```
      DO 2 I2=1,INRD2
        ABDIF(I2)=-1.0
    2 CONTINUE
      DO 51 LX=1,INRD
        R(LX)=0.0
   51 CONTINUE
C     **** GENERATES EXTR. VAL. RANDOM DEVIATES ****
      CALL GGWIB(DSEED,A,NRD,R)
      DO 1 I1=1,NRD
          R(I1)=-(LOG(B*R(I1)+C))
    1 CONTINUE
      LOGB=-(LOG(B))
      THETA=1.00/A
C     ************************************************
C     **** CALCULATES MAXIMUM LIKELIHOOD ESTIMATES ****
      SUM1=0
      SUM2=0
      SUM3=0
      SUM4=0
      DO 3 I3=1,NRD
        SUM1=R(I3)+SUM1
    3 CONTINUE
      DO 4 I4=1,MAXLIK
        DO 5 I5=1,NRD
           ZX=-(R(I5)/THETA)
           SUM2=R(I5)*EXP(ZX)+SUM2
           SUM3=EXP(ZX)+SUM3
    5   CONTINUE
      THETA1=(SUM1/NRD)-(SUM2/SUM3)
      DIF=ABS(THETA1-THETA)
      THETA=THETA1
      IF(DIF.LE.0.00001) GOTO100
    4 CONTINUE
  100 DO 6 I6=1,NRD
         ZX1=-(R(I6)/THETA)
         SUM4=EXP(ZX1)+SUM4
    6 CONTINUE
      EPS=(-THETA)*(LOG(SUM4/NRD))
C     ********************************
C     **** ORDERS FIRST NRD ELEMENTS OF ARRAY R ****
      DO 77 IOR1=1,NRD-1
        IN=IOR1+1
         DO 78 IOR2=IN,NRD
          IF(R(IOR1).LT.R(IOR2)) GOTO78
          SW=R(IOR1)
          R(IOR1)=R(IOR2)
          R(IOR2)=SW
   78   CONTINUE
   77 CONTINUE
C     **************************************************
C     **** GENERATES DNA STATISTICS ****
      SN(1)=0.0
      DO 7 I7=2,NRD+1
        XI7=REAL(I7)
```

```
        XNRD=REAL(NRD)
      IF(SNC.EQ.1) THEN
        SN(I7)=(XI7-1.0)/(XNRD)
         ELSE IF(SNC.EQ.2) THEN
          SN(I7)=((XI7-1.0)-.5)/(XNRD)
           ELSE IF(SNC.EQ.3) THEN
            SN(I7)=((XI7-1.0)-.3)/(XNRD+.4)
           ELSE IF(SNC.EQ.4) THEN
            SN(I7)=(((XI7-1.0)/(XNRD+1.0))+((XI7-1.0)/(XNRD-1.0)))/2.0
        ELSE IF(SNC.EQ.5) THEN
         SN(I7)=((XI7-1.0)-.375)/(XNRD+.25)
       ELSE IF(SNC.EQ.6) THEN
        SN(I7)=(XI7-1.0)/(XNRD+1.0)
      ENDIF
    7 CONTINUE
      DO 33 J1=1,NRD
        CDF=EXP(-EXP((EPS-R(J1))/THETA))
        ABDIF(J1)=ABS(CDF-SN(J1))
        ABDIF(J1+NRD)=ABS(CDF-SN(J1+1))
   33 CONTINUE
      CALL VSRTA(ABDIF,INRD2)
      DNA(I)=ABDIF(INRD2)
  501 CONTINUE
      CALL VSRTA(DNA,REP)
C     **********************************************
C     **** ESTIMATES ENDPOINTS, INTERPOLATES BETWEEN ****
C     ******* POINTS, AND FINDS CRITICAL VALUES *********
      DIM=REP+2
      DO 8 I8=REP,1,-1
        TDNA(I8+1)=DNA(I8)
    8 CONTINUE
      Y(1)=0.0
      Y(DIM)=1.0
      IF(COND1.EQ.1) THEN
        CALL PLOTO(Y,DIM)
        CALL ENDP(TDNA,Y,DIM,REP)
        PRINT*,' '
        CALL PRNT(REP,NRD,SNC)
        PRINT*,' PLP #1'
        CALL CRITVA(Y,TDNA,DIM)
      ENDIF
      IF(COND2.EQ.1) THEN
        CALL PLOT1(Y,DIM)
        CALL ENDP(TDNA,Y,DIM,REP)
        PRINT*,' '
        CALL PRNT(REP,NRD,SNC)
        PRINT*,' PLP#2'
        CALL CRITVA(Y,TDNA,DIM)
      ENDIF
      IF(COND3.EQ.1) THEN
        CALL PLOT2(Y,DIM)
        CALL ENDP(TDNA,Y,DIM,REP)
        PRINT*,' '
        CALL PRNT(REP,NRD,SNC)
```

```
          PRINT*,' PLP#3'
          CALL CRITVA(Y,TDNA,DIM)
        ENDIF
        IF(COND4.EQ.1) THEN
          CALL PLOT3(Y,DIM)
          CALL ENDP(TDNA,Y,DIM,REP)
          PRINT*,' '
          CALL PRNT(REP,NRD,SNC)
          PRINT*,' PLP#4'
          CALL CRITVA(Y,TDNA,DIM)
        ENDIF
        IF(COND5.EQ.1) THEN
          CALL PLOT4(Y,DIM)
          CALL ENDP(TDNA,Y,DIM,REP)
          PRINT*,' '
          CALL PRNT(REP,NRD,SNC)
          PRINT*,' PLP#5'
          CALL CRITVA(Y,TDNA,DIM)
        ENDIF
        IF(COND6.EQ.1) THEN
          CALL PLOT5(Y,DIM)
          CALL ENDP(TDNA,Y,DIM,REP)
          PRINT*,' '
          CALL PRNT(REP,NRD,SNC)
          PRINT*,' PLP#6'
          CALL CRITVA(Y,TDNA,DIM)
        ENDIF
        IF(IZZ.GT.5) GOTO301
  500 CONTINUE
  301 STOP
      END
C     ********************************
C     ********* SUBROUTINES *********
C     ********************************
C     **** PLOTS ALONG THE Y-AXIS USING PLP#1 ***
      SUBROUTINE PLOTO(Y,IDIM)
      DIMENSION Y(IDIM)
      IA=IDIM-1
      DO 9 MO=2,IA
         MO1=MO-1.0
         Y(MO)=(MO1)/(IDIM-2.0)
    9 CONTINUE
      RETURN
      END
C     ********************************************
C
C     **** PLOTS ALONG THE Y-AXIS USING PLP#2 ***
      SUBROUTINE PLOT1(Y,IDIM)
      DIMENSION Y(IDIM)
      IB=IDIM-1
      DO 10 M1=2,IB
        M11=M1-1.0
        Y(M1)=(M11-.5)/(IDIM-2.0)
   10 CONTINUE
```

```
      RETURN
      END
C     ****************************************
C
C     *** PLOTS ALONG Y-AXIS USING PLP#3 ***
      SUBROUTINE PLOT2(Y,IDIM)
      DIMENSION Y(IDIM)
      IC=IDIM-1
      DO 12 M2=2,IC
        M21=M2-1.0
        Y(M2)=(M21-.3)/((IDIM-2)+.4)
   12 CONTINUE
      RETURN
      END
C     ****************************************
C
C     *** PLOTS ALONG Y-AXIS USING PLP#4 ***
      SUBROUTINE PLOT3(Y,IDIM)
      DIMENSION Y(IDIM)
      ID=IDIM-1
      DO 14 M3=2,ID
        M31=M3-1.0
        Y(M3)=(M31/((IDIM-2.0)+1.0)+(M31-1.0)/((IDIM-2.0)-1.0))/2.0
   14 CONTINUE
      RETURN
      END
C     ****************************************
C
C     *** PLOTS ALONG Y-AXIS USING PLP#5 ***
      SUBROUTINE PLOT4(Y,IDIM)
      DIMENSION Y(IDIM)
      IE=IDIM-1
      DO 16 M4=2,IE
        M41=M4-1.0
        Y(M4)=(M41-.375)/((IDIM-2.0)+.25)
   16 CONTINUE
      RETURN
      END
C     ****************************************
C
C     *** PLOTS ALONG Y-AXIS USING PLP#6 ***
      SUBROUTINE PLOT5(Y,IDIM)
      DIMENSION Y(IDIM)
      IG=IDIM-1
      DO 18 M5=2,IG
         Y(M5)=(M5-1.0)/((IDIM-2.0)+1.0)
   18 CONTINUE
      RETURN
      END
C     ****************************************
C
C     *** ESTIMATES ENDPOINTS ***
      SUBROUTINE ENDP(X,Y,IV,IM)
      DIMENSION X(IV),Y(IV)
```

```
      SLOPE=(Y(2)-Y(3))/(X(2)-X(3))
      B=Y(2)-SLOPE*X(2)
      T1=(-B)/SLOPE
      X(1)=T1
      SLOPE=(Y(IW)-Y(IW+1))/(X(IW)-X(IW+1))
      B1=Y(IW)-SLOPE*X(IW)
      T2=(1.0-B1)/SLOPE
      X(IV)=T2
      RETURN
      END
C     ****************************
C
C     *** FINDS CRITICAL VALUES ***
      SUBROUTINE CRITVA(Y,TDNA,IDIM)
      DIMENSION Y(IDIM),TDNA(IDIM)
      NZA=IDIM-1
      DO 30 J=80,95,5
        DO 31 K=NZA,1,-1
          IF(Y(K).EQ.J/100.0) THEN
            CVAL=TDNA(K)
           PRINT*,'THE ',J,'TH%= ',CVAL
             GOTO 30
          ENDIF
          IF(Y(K).LT.J/100.0) THEN
            SLOPE=(Y(K)-Y(K+1))/(TDNA(K)-TDNA(K+1))
            Z=Y(K)-SLOPE*TDNA(K)
            CVAL=((J/100.0)-Z)/SLOPE
       PRINT*,'THE ',J,'TH%= ',CVAL
            GOTO30
      ENDIF
   31  CONTINUE
   30 CONTINUE
      DO 32 L=NZA,1,-1
        IF(Y(L).EQ.0.99) THEN
          CVAL=TDNA(L)
         PRINT*,'THE 99TH%= ',CVAL
          GOTO 42
        ENDIF
        IF(Y(L).LT.0.99) THEN
            SLOPE=(Y(L)-Y(L+1))/(TDNA(L)-TDNA(L+1))
            Z1=Y(L)-SLOPE*TDNA(L)
            CVAL=(0.99-Z1)/SLOPE
           PRINT*,'THE 99TH%= ',CVAL
            GOTO42
        ENDIF
   32 CONTINUE
   42 RETURN
      END
C     ****************************
C
C     *** PRINTS HEADINGS ***
      SUBROUTINE PRNT(JREP,NRD,ISNC)
      PRINT*,'**************************'
      PRINT*,' # OF REPS= ',JREP
```

```
PRINT*,' # OF RAND. DEV.= ',NRD
PRINT*,' SN: PLP#',ISNC
RETURN
END
```

```
DRR,T500,CM65000.  T800627,ROGERS,4458
ATTACH,IMSL,ID=LIBRARY,SN=ASD.
LIBRARY,IMSL.
FTN5,ANSI=0.
LGO.
*EOR
      PROGRAM POWER
C     *****************************************
C     * DIST=1: EXT. VAL.(TYPE1,LARGE VAL.) *
C     * DIST=2: STANDARD NORMAL             *
C     * DIST=3: WEIBULL                     *
C     * DIST=4: LOG-NORMAL                  *
C     * DIST=5: CHI-SQUARED(1 D.F.)
C     * DIST=6: CHI-SQUARED(4 D.F.)
C     *****************************************
      DIMENSION RDEV(40),RAND(41),ABDIF(80),SN(41)
      DIMENSION RDEV1(40),RDEV2(40),RDEV3(40)
      INTEGER DIST,REPS,REJ5,REJ1
      DOUBLE PRECISION DSEED,DSEED1
      DOUBLE PRECISION DSEED2,DSEED3
      DATA A,B,C,D,E/2*1.0,2*0.0,1.0/
      DIST=3
      IREPS=250
      PRINT*,' REPS=',IREPS
      PRINT*,' ++ DISTRIBUTION #',DIST,' ++'
      DSEED=1234567.0D0
      DSEED1=4567123.0D0
      DSEED2=3451267.0D0
      DSEED3=2347651.0D0
      NRD=40
      NRD2=NRD*2
      REJ5=0
      REJ1=0
      DO 98 REPS=1,2500
C     *********************
C     * GENERATES DIST.#1 *
C     *********************
      IF(DIST.EQ.1) THEN
        CALL GGWIB(DSEED,A,NRD,RDEV)
           DO1 I1=1,NRD
             RDEV(I1)=-(LOG(B*RDEV(I1)+C))
    1      CONTINUE
      ENDIF
C     *********************
C     * GENERATES DIST.#2 *
C     *********************
      IF(DIST.EQ.2) THEN
        CALL GGNML(DSEED,NRD,RDEV)
      ENDIF
C     *********************
C     * GENERATES DIST.#3 *
C     *********************
      IF(DIST.EQ.3) THEN
        CALL GGWIB(DSEED,A,NRD,RDEV)
```

```
      ENDIF
C     **********************
C     * GENERATES DIST.#4 *
C     **********************
      IF(DIST.EQ.4) THEN
        CALL GGNLG(DSEED,NRD,D,E,RAND)
          DO99 J=1,NRD
            RDEV(J)=RAND(J)
   99     CONTINUE
      ENDIF
C     **********************
C     * GENERATES DIST#5 *
C     *********************
      IF(DIST.EQ.5) THEN
        CALL GGNML(DSEED,NRD,RDEV)
          DO22 I22=1,NRD
            RDEV(I22)=RDEV(I22)*RDEV(I22)
   22   CONTINUE
      ENDIF
C     ********************
C     * GENERATES DIST#6 *
C     ********************
      IF(DIST.EQ.6) THEN
        CALL GGNML(DSEED,NRD,RDEV)
        CALL GGNML(DSEED1,NRD,RDEV1)
        CALL GGNML(DSEED2,NRD,RDEV2)
        CALL GGNML(DSEED3,NRD,RDEV3)
          DO9 I9=1,NRD
      RDEV(I9)=RDEV(I9)**2+RDEV1(I9)**2+RDEV2(I9)**2+RDEV3(I9)**2
    9 CONTINUE
      ENDIF
C     ***********************************
C     * CALCULATES PARAMETER ESTIMATES *
C     ***********************************
      EPS=0.0
      THETA=1.0
      SUM1=0.0
      SUM2=0.0
      SUM3=0.0
      SUM4=0.0
      DO3 I3=1,NRD
        SUM1=RDEV(I3)+SUM1
    3 CONTINUE
      DO4 I4=1,1000
        DO5 I5=1,NRD
          ZX=-(RDEV(I5)/THETA)
          SUM2=RDEV(I5)*EXP(ZX)+SUM2
          SUM3=EXP(ZX)+SUM3
    5   CONTINUE
      THETA1=(SUM1/NRD)-(SUM2/SUM3)
      DIF=ABS(THETA1-THETA)
      THETA=THETA1
      IF(DIF.LE.0.00001) GOTO 100
    4 CONTINUE
```

```
100 DO6 I6=1,NRD
      ZX1=-(RDEV(I6)/THETA)
      SUM4=EXP(ZX1)+SUM4
  6 CONTINUE
    EPS=(-THETA)*(LOG(SUM4/NRD))
C     ************************************************
C     * ORDERS RDEV, FINDS DMAX, COUNTS REJECTIONS *
C     ************************************************
    CALL VSRTA(RDEV,NRD)
    SN(1)=0.0
    DO7 I7=2,NRD+1
      XI7=REAL(I7)
      XNRD=REAL(NRD)
    SN(I7)=(XI7-1.0)/(XNRD)
  7 CONTINUE
    DO50 ID=1,NRD
    CDF=EXP(-EXP((EPS-RDEV(ID))/THETA))
      ABDIF(ID)=ABS(CDF-SN(ID))
    ABDIF(ID+NRD)=ABS(CDF-SN(ID+1))
 50 CONTINUE
    CALL VSRTA(ABDIF,NRD2)
    DMAX=ABDIF(NRD2)
    IF(DMAX.GT.0.1344) REJ5=REJ5+1
    IF(DMAX.GT.0.1499) REJ1=REJ1+1
 98 CONTINUE
    XREJ5=REAL(REJ5)
    XREJ1=REAL(REJ1)
    PERC5=XREJ5/2500.0
    PERC1=XREJ1/2500.0
    PRINT*,'*****************'
    PRINT*,'  NRD= ',NRD
    PRINT*,'*****************'
    PRINT*,' REJ5= ',REJ5,'  REJ1= ',REJ1
    PRINT*,'**********'
    PRINT*,'  -SN1- '
    PRINT*,' ALPHA=.05'
    PRINT*,'% REJECTED= ',PERC5
    PRINT*,' '
    PRINT*,' ALPHA=.01'
    PRINT*,'% REJECTED =',PERC1
    PRINT*,' '
    PRINT*,'**********'
    STOP
    END
```

```
DRR,T400,CM65000. T800627,ROGERS,4458
ATTACH,IMSL,ID=LIBRARY,SN=ASD.
LIBRARY,IMSL.
FTN5,ANSI=0.
LGO.
*EOR
      PROGRAM POWERA
C     ****************************************
C     * DIST=1: EXT. VAL.(TYPE1,LARGE VAL.) *
C     * DIST=2: STANDARD NORMAL             *
C     * DIST=3: WEIBULL                     *
C     * DIST=4: LOG-NORMAL                  *
C     * DIST=5: CHI-SQUARED(1 D.F.)         *
C     * DIST=6: CHI-SQUARED(4 D.F.)         *
C     ****************************************
      DIMENSION RDEV(40),RAND(41),ABDIF3(80),ABDIF5(80),ABDIF6(80)
      DIMENSION SN3(41),SN5(41),SN6(41)
      DIMENSION RDEV1(40),RDEV2(40),RDEV3(40)
      INTEGER DIST,REPS,REJ35,REJ31,REJ55,REJ51,REJ65,REJ61
      DOUBLE PRECISION DSEED,DSEED1
      DOUBLE PRECISION DSEED2,DSEED3
      DATA A,B,C,D,E/2*1.0,2*0.0,1.0/
      DIST=1
      PRINT*,' ++ DISTRIBUTION #',DIST,' ++'
      DSEED=1234567.0D0
      DSEED1=4567123.0D0
      DSEED2=3451267.0D0
      DSEED3=2347651.0D0
      NRD=40
      NRD2=NRD*2
      REJ35=0
      REJ55=0
      REJ31=0
      REJ51=0
      REJ65=0
      REJ61=0
      DO 98 REPS=1,2500
C     **********************
C     * GENERATES DIST.#1 *
C     **********************
      IF(DIST.EQ.1) THEN
        CALL GGWIB(DSEED,A,NRD,RDEV)
           DO1 I1=1,NRD
             RDEV(I1)=-(LOG(B*RDEV(I1)+C))
    1      CONTINUE
      ENDIF
C     **********************
C     * GENERATES DIST.#2 *
C     **********************
      IF(DIST.EQ.2) THEN
        CALL GGNML(DSEED,NRD,RDEV)
      ENDIF
```

```
C     *********************
C     * GENERATES DIST.#3 *
C     *********************
      IF(DIST.EQ.3) THEN
        CALL GGNIB(DSEED,A,NRD,RDEV)
      ENDIF
C     *********************
C     * GENERATES DIST.#4 *
C     *********************
      IF(DIST.EQ.4) THEN
        CALL GGNLG(DSEED,NRD,D,E,RAND)
          DO99 J=1,NRD
            RDEV(J)=RAND(J)
   99     CONTINUE
      ENDIF
C     ***********************
C     * GENERATES DIST.#5 *
C     *********************
      IF(DIST.EQ.5) THEN
        CALL GGNML(DSEED,NRD,RDEV)
          DO22 I22=1,NRD
            RDEV(I22)=RDEV(I22)*RDEV(I22)
   22     CONTINUE
      ENDIF
C     *********************
C     * GENERATES DIST.#6 *
C     *********************
      IF(DIST.EQ.6) THEN
        CALL GGNML(DSEED,NRD,RDEV)
        CALL GGNML(DSEED1,NRD,RDEV1)
        CALL GGNML(DSEED2,NRD,RDEV2)
        CALL GGNML(DSEED3,NRD,RDEV3)
          DO9 I9=1,NRD
            RDEV(I9)=RDEV(I9)**2+RDEV1(I9)**2+RDEV2(I9)**2+RDEV3(I9)**2
    9       CONTINUE
ENDIF
C     ***********************************
C     * CALCULATES PARAMETER ESTIMATES *
C     ***********************************
      EPS=0.0
      THETA=1.0
      SUM1=0.0
      SUM2=0.0
      SUM3=0.0
      SUM4=0.0
      DO3 I3=1,NRD
        SUM1=RDEV(I3)+SUM1
    3 CONTINUE
      DO4 I4=1,1000
        DO5 I5=1,NRD
          ZX=-(RDEV(I5)/THETA)
          SUM2=RDEV(I5)*EXP(ZX)+SUM2
          SUM3=EXP(ZX)+SUM3
    5   CONTINUE
```

```
      THETA1=(SUM1/NRD)-(SUM2/SUM3)
      DIF=ABS(THETA1-THETA)
      THETA=THETA1
      IF(DIF.LE.0.00001) GOTO 100
    4 CONTINUE
  100 DO6 I6=1,NRD
        ZX1=-(RDEV(I6)/THETA)
        SUM4=EXP(ZX1)+SUM4
    6 CONTINUE
      EPS=(-THETA)*(LOG(SUM4/NRD))
C     ***********************************************
C     * ORDERS RDEV, FINDS DMAX, COUNTS REJECTIONS *
C     ***********************************************
      CALL VSRTA(RDEV,NRD)
      SN3(1)=0.0
      SN5(1)=0.0
      SN6(1)=0.0
      DO7 I7=2,NRD+1
        XI7=REAL(I7)
        XNRD=REAL(NRD)
        SN3(I7)=((XI7-1.0)-.3)/(XNRD+.4)
        SN5(I7)=((XI7-1.0)-.375)/(XNRD+.25)
        SN6(I7)=(XI7-1.0)/(XNRD+1.0)
    7 CONTINUE
      DO50 ID=1,NRD
      CDF=EXP(-EXP((EPS-RDEV(ID))/THETA))
        ABDIF3(ID)=ABS(CDF-SN3(ID))
        ABDIF3(ID+NRD)=ABS(CDF-SN3(ID+1))
        ABDIF5(ID)=ABS(CDF-SN5(ID))
        ABDIF5(ID+NRD)=ABS(CDF-SN5(ID+1))
        ABDIF6(ID)=ABS(CDF-SN6(ID))
        ABDIF6(ID)=ABS(CDF-SN6(ID+1))
   50 CONTINUE
      CALL VSRTA(ABDIF3,NRD2)
      CALL VSRTA(ABDIF5,NRD2)
      CALL VSRTA(ABDIF6,NRD2)
      DMAX3=ABDIF3(NRD2)
      DMAX5=ABDIF5(NRD2)
      DMAX6=ABDIF6(NRD2)
      IF(DMAX3.GT.0.1407) REJ35=REJ35+1
      IF(DMAX3.GT.0.1659) REJ31=REJ31+1
      IF(DMAX5.GT.0.1405) REJ55=REJ55+1
      IF(DMAX5.GT.0.1657) REJ51=REJ51+1
      IF(DMAX6.GT.0.1404) REJ65=REJ65+1
      IF(DMAX6.GT.0.1637) REJ61=REJ61+1
   98 CONTINUE
      XREJ35=REAL(REJ35)
      XREJ31=REAL(REJ31)
      XREJ55=REAL(REJ55)
      XREJ51=REAL(REJ51)
      XREJ65=REAL(REJ65)
      XREJ61=REAL(REJ61)
      PRINT*,'****************'
      P35=XREJ35/2500.0
```

```
P31=XREJ31/2500.0
P55=XREJ55/2500.0
P51=XREJ51/2500.0
P65=XREJ65/2500.0
P61=XREJ61/2500.0
PRINT*,'  NRD= ',NRD
PRINT*,'****************'
PRINT*,' '
PRINT*,'**********'
PRINT*,'  -SN3- '
PRINT*,' ALPHA=.05'
PRINT*,' % REJECTED= ',P35
PRINT*,' '
PRINT*,' ALPHA=.01'
PRINT*,' % REJECTED= ',P31
PRINT*,' '
PRINT*,'**********'
PRINT*,'  -SN5-'
PRINT*,' ALPHA=.05'
PRINT*,' % REJECTED= ',P55
PRINT*,' '
PRINT*,' ALPHA=.01'
PRINT*,' % REJECTED= ',P51
PRINT*,' '
PRINT*,'**********'
PRINT*,'  -SN6-'
PRINT*,' ALPHA=.05'
PRINT*,' % REJECTED= ',P65
PRINT*,' '
PRINT*,' ALPHA=.01'
PRINT*,' % REJECTED= ',P61
STOP
END
```

VITA

Douglas R. Rogers was born on 28 June 1959 in Denver, Colorado to David and Arlene Rogers. He graduated from Summerville High School, Summerville, South Carolina in 1977 and received a Bachelor of Science degree in Mathematics from Baptist College at Charleston. In May 1980, he received a commission in the United States Air Force through the AFROTC program. Lieutenant Rogers entered the Air Force Institute of Technology Resident School of Engineering in June 1980.

Permanent address: 231 Tupperway Rd.
Summerville, South Carolina
29483

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/GOR/MA/81D-11	2. GOVT ACCESSION NO. AD-A115 544	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A MONTE CARLO TECHNIQUE USING MODIFIED LINEAR INTERPOLATION TO GENERATE KOLMOGOROV-SMIRNOV CRITICAL VALUES FOR THE EXTREME VALUE DISTRIBUTION		5. TYPE OF REPORT & PERIOD COVERED MS Thesis 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Douglas R. Rogers 2Lt USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AFIT-EN) Wright-Patterson AFB, Ohio 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE December, 1981 13. NUMBER OF PAGES 106
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

15 APR 1982

18. SUPPLEMENTARY NOTES

Approved for public release, AFR 190-17

~~FREDERICK LYNCH, Major, USAF~~
~~Director of Public Affairs~~

Lynn E. Wolaver

Dean for Research and Professional Development
Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Extreme Value Distribution
Linear Interpolation
Critical Value Estimation
Monte Carlo Simulation
Plotting Positions
Kolmogorov-Smirnov Test
Goodness-of-Fit Test

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An investigation was conducted to examine the merits of an estimation technique involving linear interpolation to estimate Kolmogorov-Smirnov (K-S) critical values when the scale and location parameters of the hypothesized distribution are unknown. The purpose of the linear estimation technique is to reduce useful critical values for the K-S goodness-of-fit test.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Also, different plotting positions were studied to ascertain which plotting positions used in calculating and plotting the K-S test statistic values provided the best critical value estimation.

The distribution used as the hypothesized distribution was the Type I, extreme value distribution (largest extreme value).

In addition, a power study was performed which compared the power of the true critical values against the power of the estimated critical values.

The following are the major results. Useful critical values were found with the linear estimation technique using relatively few Monte Carlo generated samples. Further, the plotting positions found to be the best in calculating the K-S test statistic values and plotting these values were respectively i/n and (i-.5)/n where i is the ith ranked point in a sample of size n.

DATE
FILMED

-8